



Le dictionnaire électronique des séquences nominales figées en coréen et de leurs formes fléchies - méthodes et applications

Sun-Mee Bae

► To cite this version:

Sun-Mee Bae. Le dictionnaire électronique des séquences nominales figées en coréen et de leurs formes fléchies - méthodes et applications. Autre [cs.OH]. Université Paris-Est, 2002. Français. NNT : . tel-00627610

HAL Id: tel-00627610

<https://theses.hal.science/tel-00627610>

Submitted on 29 Sep 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université de Marne-la-Vallée
Institut Gaspard Monge

**Le dictionnaire électronique des séquences nominales figées
en coréen et de leurs formes fléchies**

- méthodes et applications -

Sun-Mee Bae

Thèse de doctorat en Informatique fondamentale
Soutenue le 25 septembre 2002

Directeur de thèse : **Eric Laporte**

Jury : **Maxime Crochemore**
Gaston Gross (*rapporteur, président*)
Hyun-Kwon Kim
Eric Laporte
Denis Maurel (*rapporteur*)

REMERCIEMENTS

D'abord, je tiens à remercier tous les membres de jury d'avoir accepté de participer à la soutenance.

Je tiens à exprimer ma gratitude envers mes collègues du LADL et ceux du IGM. Ils m'ont beaucoup aidé dans mes travaux, et grâce à eux j'ai pu passer la période de thèse dans une atmosphère très amicale. Surtout, je n'oublierai pas les relations chaleureuses et amicales avec Christian Leclère, Blandine Coutois, Annie Meunier, Max Silberztein, Krit Kosawat, Matthieu Constant, Takuya Nakamura, Sébastien Paumier, Marta Ungermanova et tous mes collègues coréens d'IGM. Les autres membres et visiteurs du laboratoire m'ont aussi beaucoup encouragé tout au long de mon travail, et je les en remercie.

J'ai une dette particulière envers mes amis du séminaire pour la linguistique coréenne : Go Yeon-chan, Kim Jong-duk, Soe Jeong-yeon, et Yoon Hui-seon qui ont pensé ensemble à mes problèmes scientifiques comme leurs problèmes et qui m'ont donné de précieux conseils.

Je tiens à exprimer ma reconnaissance envers mon inoubliable directeur de thèse, le Professeur Eric Laporte, sans qui je n'aurais pas pu imaginer ma thèse. Il m'a permis de m'initier à la linguistique informatique, il m'a accordé sa confiance et a été présent à toutes les étapes. Surtout, sa passion pour les sciences m'a poussé sans arrêt à travailler avec une attitude très sérieuse et scientifique.

Enfin, je suis reconnaissante envers toute ma famille à commencer par mon mari et mon fils qui m'ont attendu longtemps en m'encourageant.

TABLE DES MATIÈRES

REMERCIEMENTS.....	i
NOTATIONS.....	v
INTRODUCTION.....	1
 Chapitre 1. Plusieurs points de vue sur les noms composés du coréen.....	 6
1. Les noms composés du point de vue de la formation des mots.....	6
1-1. Les noms simples, les noms dérivés et les noms composés.....	6
1-2. Exemples proprement coréens.....	9
1-3. La formation des mots sino-coréens.....	14
2. Les noms composés du point de vue du traitement informatique.....	18
2-1. Problèmes spécifiques du traitement des noms composés.....	18
2-2. Travaux apparentés sur les noms composés.....	20
2-3. Discussion.....	25
3. Deux dictionnaires électroniques des noms fondés sur la formation des mots.....	26
4. Conclusion.....	33
 Chapitre 2. Les séquences nominales figées dans un dictionnaire électronique..	 35
1. La notion de figement : les séquences nominales figées et libres.....	35
2. Les critères de figement.....	37
2-1. Opacité sémantique.....	38
2-2. Restriction distributionnelle.....	38
2-3. Paramètres syntaxiques.....	38
2-3-1. Blocage des propriétés transformationnelles.....	38
2-3-2. Restriction sur l'insertion.....	40
2-3-3. Acceptabilité de N_1N_2 est (un N_2 + un type de N_2)	41
2-4. Les séquences nominales figées de la structure XN	42
2-4-1. La définition de pseudo-mot X	43
2-4-2. Plusieurs types de XN	49
3. Variations typographiques des séquences nominales figées.....	58
3-1. Différences formelles entre les composés français et coréens.....	58
3-2. Variations typographiques.....	60

3-2-1. Désaccords entre les dictionnaires éditoriaux.....	60
3-2-2. Désaccords entre les dictionnaires et les textes.....	61
3-3. Classification des séquences nominales figées en fonction des conventions typographiques.....	63
4. Conclusion.....	67

Chapitre 3. Description des séquences nominales figées dans un dictionnaire électronique..... 68

1. Espacement.....	68
2. Informations flexionnelles.....	69
2-1. La combinatoire entre les noms et les postpositions nominales.....	69
2-2. Classification des séquences nominales figées selon les contraintes combinatoires.....	72
3. Origine.....	73
4. Catégorie grammaticale des éléments.....	76
5. Les informations sémantiques.....	77
5-1. Noms humains, noms d'animaux et noms de plantes.....	77
5-1-1. Compatibilité avec la Postp.e.	78
5-1-2. Co-occurrence avec les noms classifieurs.....	79
5-2. Les humains collectifs et les humains institutionnels.....	80
6. La fonction de sous-classificateur de N_1 et X dans N_1N_2 et XN	81
7. Les informations syntaxiques.....	82
7-1. Définitions.....	82
7-2. Les noms compacts prédicatifs ($Npred$)	84
7-2-1. $Npred (= : V-n)$	84
7-2-2. Les $Npred$ qui ne sont pas des $V-n$	88
7-3. Les noms figés prédicatifs ($NFpred$)	90
7-3-1. $NFpred (= : V-n)$	90
7-3-2. $NFpred$ à nom tête déverbal.....	91
7-3-3. Les $NFpred$ qui ne sont pas des $V-n$	93
7-4. Quelques remarques de la relation de paraphrase dans la structure $X Npred$	95
7-4-1. Type A: $Rx (-jeok + *-han + E) Npred$	95
7-4-2. Type B: $Rx (*-jeok + -han + E) Npred$	98
8. Conclusion.....	99

Chapitre 4. Dictionnaires représentés par des graphes : séquences nominales figées

et séquences de postpositions nominales.....	102
1. Choix entre la représentation sous forme de liste ou de graphes.....	102
2. Dictionnaire de séquences nominales représentées par des graphes.....	103
2-1. Un exemple de graphes de séquences nominales à espacement facultatif.....	104
2-2. Un exemple de graphes de séquences nominales compactes et à espacement facultatif.....	105
3. Dictionnaire de séquences de postpositions nominales représentées par des graphes.....	107
3-1. Classification linguistique des postpositions nominales.....	107
3-1-1. Critère syntaxique.....	107
3-1-2. Critères distributionnels.....	110
3-2. Le marqueur pluriel <i>-deul</i> en coréen.....	112
3-2-1. Problématique du marqueur pluriel <i>-deul</i>	112
3-2-2. Restrictions distributionnelles sur <i>-deul</i>	116
3-2-3. Deux emplois de <i>-deul</i> : pluriel du substantif et pluriel du sujet.....	117
3-2-4. Relations de paraphrase à <i>-deul</i> qui marque le pluriel du sujet de la phrase...	129
3-3. Combinatoire des postpositions nominales.....	134
3-3-1. Forme canonique, variantes phonologiquement conditionnées, variantes libres et homonymes.....	134
3-3-2. Combinatoire des postpositions nominales.....	136
4. Conclusion.....	143

Chapitre 5. Applications des dictionnaires électroniques de séquences

nominales figées et de leurs formes fléchies.....	144
1. Construction automatique du dictionnaire des formes fléchies des séquences nominales figées avec INTEX.....	144
1-1. Génération automatique des formes fléchies des séquences nominales figées.....	144
1-2. Résultats.....	148
1-3. Compression du DELAF et du DELACF de NFF en FST déterministes minimaux.....	151
2. Reconnaissance automatique des NFF par dictionnaire électronique.....	158
2-1. Reconnaissance des NFF avec INTEX.....	158
2-2. Perspectives pour le traitement des séquences nominales non reconnues.....	162
3. Avantages de notre dictionnaire électronique pour plusieurs applications du traitement automatique.....	165

3-1. Recherche d'informations et extraction de mots-clés.....	165
3-2. Analyse syntaxique.....	168
3-3. Traduction automatique.....	171
4. Conclusion.....	172
CONCLUSION	173
BIBLIOGRAPHIE	175
ANNEXE 1. Définition de l'alphabet du coréen dans INTEX sous Windows 98 ou 2000 en coréen.....	183
ANNEXE 2. Extrait du dictionnaire des NFF.....	189
ANNEXE 3. Extrait du dictionnaire des séquences de postpositions nominales.....	191

NOTATIONS

^	Espacement facultatif
-	frontière des mots et des éléments fonctionnels
A	Adjectif
Acc	Postposition de l'accusatif
Adv	Adverbe
Aux	Postposition auxiliaire
Conj	Postposition conjonctive
Copu	Postposition du copulatif
Dét	Déterminant
E	Mot vide
Emp	Élément emprunté à d'autres langues que le chinois
Gén	Postposition du génitif
Mmod	Morphème de modalité
Mpas	Morphème de passé
Mplu	Morphème de pluriel
N	Nom
N₀	Sujet
Ncomp	Nom complémenteur
Ni	Nom incomplet
Nmtf	Postposition du nominatif
Pc	Élément proprement coréen
Pf	Préfixe
Postp	Postposition d'un complément non accusatif
Postp.buteo	Postposition du complément en <i>buteo</i>
Postp.e	Postposition du complément en <i>e</i>
Postp.e daeha	Postposition du complément en <i>e daeha</i>
Postp.ege	Postposition du complément en <i>ege</i>
Postp.eseo	Postposition du complément en <i>eseo</i>
Postp.eulo	Postposition du complément en <i>eulo</i>
Postp.hago	Postposition du complément en <i>hago</i>
Postp.lo	Postposition du complément en <i>lo</i>
Postp.loseo	Postposition du complément en <i>loseo</i>
Ra	Radical adjectival
Rv	Radical verbal

Rx	Radical sans catégorie grammaticale
Sc	Élément sino-coréen
Scomp	Suffixe complémenteur
Sconj	Suffixe verbal conjonctif
Sf	Suffixe
Sfa	Suffixe de dérivation adjectivale
Sfd	Suffixe verbal déterminatif
Sflx	Suffixe de flexion
Sfn	Suffixe de nominalisation
Sfv	Suffixe de verbalisation
St	Suffixe terminal du mode déclaratif
Sti	Suffixe terminal du mode interrogatif
StImp	Suffixe terminal du mode impératif
V	Verbe
Vaux	Verbe auxiliaire
V-n	Substantif morpho-syntaxiquement associé à un verbe
Voc	Postposition du vocatif
X	Élément sans catégorie grammaticale (Pseudo-mot)
W	Séquence éventuelle de compléments

INTRODUCTION

Ce travail vise à présenter des méthodes de construction de dictionnaires électroniques de séquences nominales figées du coréen et de leurs formes fléchies, et à justifier leur validité en appliquant notre dictionnaire des séquences nominales figées à des textes coréens et en explorant les avantages dans les domaines appliqués de l'analyse automatique de textes coréens.

En tant que cadre informatique pour la théorie linguistique du LADL (Laboratoire d'Automatique Documentaire et Linguistique), les mots composés en français ou anglais sont définis comme les unités lexicales constituées de plus de deux mots simples, et ils contiennent donc au moins un séparateur (Silberztein, 1997). La distinction entre les mots simples et les mots composés n'est qu'orthographique. Par exemple, le mot *deltaplane* est un nom simple, alors que ses variantes orthographiques *delta-plane* ou *delta plane* sont des noms composés. Or, la définition ci-dessus n'est pas exactement convenable pour le nom composé du coréen, puisque l'espacement est variable selon le type de séquences nominales figées.

Dans le traitement des groupes nominaux en coréen, on est confronté à trois sortes d'agglutinations : entre noms, entre postpositions nominales, et entre noms et postpositions nominales. Par exemple :

- a. [gongsang gwahak yeonghwa]-[eseo-jocha-do]
[공상 과학 영화]-에서조차도
[N ^ N ^ N]-[Postp-Postp-Postp]
[imagination science film]-[dans-même-aussi]
= même dans les films de science-fiction

Dans (a), les trois formes *-eseo*, *-jocha*, *-do* sont des postpositions nominales (*Postp*); « ^ » indique l'espacement facultatif et « - » les frontières des mots et de chaque élément fonctionnel. Toutes les unités séparées par un tiret dans les exemples s'écrivent en fait soudées dans les textes coréens. La séquence (a) peut se réaliser graphiquement de quatre façons :

- b. gongsang gwahak yeonghwaeseojochado 공상 과학 영화에서조차도
[N N N]-[Postp Postp Postp]
gongsang gwahakyeonghwaeseojochado 공상 과학영화에서조차도
[N NN]-[Postp Postp Postp]

gongsanggwahak yeonghwaeseojochado 공상과학 영화에서조차도

[*NN* *N*]-[*Postp Postp Postp*]

gongsanggwahakyeonghwaeseojochado 공상과학영화에서조차도

[*NNN*]-[*Postp Postp Postp*]

L'espacement entre noms est irrégulier, et on ne laisse pas de blanc entre postpositions nominales ni entre noms et postpositions nominales.

D'une part, pour la reconnaissance automatique des séquences nominales figées à espacement irrégulier, il nous faudra un dictionnaire électronique qui précise leurs variations typographiques. D'autre part, pour l'analyse automatique des séquences de postpositions nominales soudées, il est indispensable de disposer d'un dictionnaire électronique des séquences de postpositions nominales qui précise la segmentation, par exemple, à l'aide d'un transducteur fini. Puisqu'en général, les formes fléchies des séquences nominales figées comme (a) apparaissent dans des textes coréens, en vue de la constitution de dictionnaires exploitables des séquences nominales figées dans des applications informatiques, il nous faudra un dictionnaire électronique des formes fléchies des séquences nominales figées. A partir de ces deux dictionnaires, à savoir des séquences nominales figées et des séquences de postpositions nominales, nous pouvons construire automatiquement le dictionnaire électronique des formes fléchies des séquences nominales figées à l'aide d'INTEX.

En fait, en coréen, les séquences composées de type NN^+ , c'est-à-dire *NN*, *NNN*, *NNNN*..., constituent la grande majorité des composés nominaux. Les combinaisons des noms sont plus libres en coréen qu'en français et sont en nombre théoriquement infini. À notre avis, il est indispensable de distinguer les séquences nominales figées des séquences nominales libres et de lister les séquences nominales figées, puisqu'il est pratiquement impossible de dresser la liste de toutes les séquences nominales dans les dictionnaires à cause de leur énorme productivité. En fonction de leur degré de figement, les séquences de noms forment un continuum entre les séquences nominales figées et les séquences nominales libres. L'exemple suivant illustre la différence entre elles :

c. [*hucheonseong* ^ *myeonyeok* ^ *gyeolpipjeung*] ^ *yebang* (*NNNN*)

[후천성 ^ 면역 ^ 결핍증] ^ 예방

« qualité acquise » « immunité » « syndrome de déficience » « prévention »

= « prévention du *SIDA* (Syndrome Immuno-Déficitaire Acquis) »

Cette séquence est une séquence nominale libre constituée de quatre noms, mais elle

comporte une séquence nominale figée. *hucheonseong myeonyeok gyeolpipjeung* « sida » constituée de trois noms qui, dans cette combinaison, n'ont pas le statut de trois unités lexicales, mais d'une seule unité avec un caractère terminologique, institutionnalisée et dont le sens n'est pas tout à fait transparent. En revanche, la combinaison entre *hucheonseong myeonyeok gyeolpipjeung* « sida » et *yebang* « prévention » obéit aux règles ordinaires de la formation de groupes nominaux du coréen : $N (= : hucheonseong myeonyeok gyeolpipjeung) - Acc \ Npred (= : yebang) - hada$ (ici, *Acc* indique la postposition de l'accusatif et *Npred* le nom prédicatif). Lors du traitement automatique des séquences nominales, outre la reconnaissance des séquences nominales figées, nous sommes confrontés à un autre type de problème, à savoir la reconnaissance des séquences nominales libres soudées. Ces deux problèmes, la reconnaissance des séquences nominales figées et la reconnaissance des séquences nominales libres soudées sont liés aux conventions typographiques. Pour réduire le nombre des séquences nominales non reconnues, il faut obligatoirement appliquer une procédure de segmentation à l'aide, notamment, des dictionnaires de séquences nominales figées. La segmentation correcte des séquences soudées du type de NN^+ du coréen est un processus indispensable pour diverses applications comme la traduction automatique, la correction d'espacement, l'indexation automatique, l'extraction de mots-clés, etc. Pour obtenir une segmentation précise des séquences nominales libres, a priori, il faut construire un dictionnaire électronique suffisamment exhaustif des séquences nominales figées, car une séquence nominale libre peut contenir une séquence nominale figée comme l'exemple (c).

Comme nous l'avons vu ci-dessus, la construction du dictionnaire électronique de séquences nominales figées est une demande inévitable dans le traitement automatique des séquences nominales du coréen en plusieurs raisons. Mais devant la construction du dictionnaire électronique, les difficultés viennent du fait que les relations entre les notions linguistiques et typographiques sont complexes : les combinaisons libres dans lesquelles entrent les mots coréens (surtout les combinaisons de mots sino-coréens) et l'espacement irrégulier posent de grandes difficultés dans la délimitation non seulement des séquences nominales figées vis-à-vis des séquences nominales libres mais aussi entre les différentes catégories de séquences nominales figées. Lors de la construction du dictionnaire électronique du coréen, nous devons faire des choix spécifiques pour appliquer la méthodologie du LADL.

Pour le dictionnaire des séquences nominales figées, nous proposons trois étapes suivantes :

- (i) Première étape : établir les critères de figement pour distinguer les séquences nominales figées des séquences nominales libres
- (ii) Deuxième étape : examiner les variations typographiques des séquences nominales figées dans des textes et des dictionnaires éditoriaux et classifier les séquences nominales figées en fonction des conventions typographiques
- (iii) Troisième étape : examiner les informations utiles à décrire dans un dictionnaire électronique pour le traitement automatique des textes coréens et les codifier pour chaque classe des séquences nominales figées

Pour le dictionnaire des séquences de postpositions nominales sous forme de graphes :

- (i) Première étape : classifier linguistiquement les postpositions nominales avec les critères formels
- (ii) Deuxième étape : considérer les formes canoniques, les variantes phonologiquement conditionnées ou libres et les homonymes, et sélectionner la forme canonique des variantes
- (iii) Troisième étape : décrire les informations sur la forme canonique et la classe linguistique des postpositions nominales à l'aide d'un transducteur fini

En vue de la constitution de dictionnaires exploitables des séquences nominales figées, nous proposons de construire le dictionnaire des formes fléchies des séquences nominales figées :

- (i) Première étape : générer automatiquement toutes les formes fléchies des séquences nominales figées à partir du code flexionnel attaché à chaque entrée et de la fonctionnalité de flexion du menu DELA d'INTEX (« Inflection ») en fusionnant les deux dictionnaires : le dictionnaire des séquences nominales et celui de séquences de postpositions nominales
- (ii) Deuxième étape : comprimer le dictionnaire des formes fléchies des séquences nominales figées en FST déterministes minimaux

Dans ce travail, nous ne comptons que les séquences nominales figées communes et nous excluons les séquences nominales propres, techniques, scientifiques et archaïques. Mais nous répertorions les noms techniques assez courants comme *hucheonseong myeonyeok gyeolpipjeung* « sida » dans notre dictionnaire. Nous avons recensé 45000 séquences nominales figées à espacement facultatif à partir des entrées

ainsi que des exemples des dictionnaires éditoriaux parus depuis 1995 : *Keumsungpan Gukeo Daesajeon* (Grand Dictionnaire du Coréen dans la version Keum-sung Publ.Co., 1997) de Keum-sung Publ.Co., *Yonsei Hangukeo Sajeon* (Yonsei Dictionnaire du Coréen, 1998) de Centre de recherches sur le coréen de l'Université Yonsei, et *Pyojun Gukeo Daesajeon* (Grand Dictionnaire Standard du Coréen, 1999) de l'Académie National de la Langue Coréen. Les sept informations fondamentales suivantes sont systématiquement décrites pour les 45000 noms figés à espacement facultatif dans notre dictionnaire électronique : l'espacement, les informations flexionnelles, la catégorie grammaticale des éléments, l'origine, les informations sémantiques, les informations syntaxiques, et les informations morpho-syntaxiques comme le statut de nom prédicatif.

Cette thèse est organisée en 5 parties.

Le chapitre 1 introduit plusieurs points de vue sur les noms composés en coréen : celui de la formation des mots, celui du traitement informatique et celui de deux dictionnaires électroniques des noms fondés sur la formation des mots.

Le chapitre 2 concerne les critères du figement pour délimiter les séquences nominales figées, puis la classification des séquences nominales figées selon les conventions typographiques.

Le chapitre 3 sera consacré aux informations fondamentales dans un dictionnaire électronique des séquences nominales figées.

Le chapitre 4 présente le dictionnaire électronique représenté par des graphes pour d'autres types des séquences nominales assez factorisables et pour les séquences des postpositions nominales.

Le chapitre 5 sera consacré à présenter une méthode de génération des formes fléchies des séquences nominales figées et à examiner les avantages de notre dictionnaire en l'appliquant dans plusieurs domaines du traitement automatique.

Chapitre 1. Plusieurs points de vue sur les noms composés du coréen

Ce chapitre est destiné à introduire plusieurs points de vue principaux sur les noms composés en coréen. Il est une étape préalable à la délimitation de la notion de nom composé dans un dictionnaire électronique. Dans la section 1, nous présentons les noms composés au niveau de la formation des mots et dans la section 2 au point de vue du traitement automatique. Enfin, la section 3 est consacrée à deux perspectives représentatives sur le dictionnaire électronique des substantifs du coréen qui sont fondamentalement basées sur la formation des mots dans le prolongement de la tradition linguistique.

1. Les noms composés du point de vue de la formation des mots

1-1. Les noms simples, les noms dérivés et les noms composés

Dans la tradition linguistique de la grammaire coréenne, l'étude des noms est focalisée sur la formation des noms. La formation des mots résulte des jeux entre les racines et les affixes.

Avant de définir les éléments de la formation des mots, nous allons commencer par introduire la notion de morphème autonome et non autonome. Au niveau morphologique, si un morphème peut constituer un mot à lui seul, il est appelé *morphème autonome*. Sinon, il est appelé *morphème non autonome*¹. Par exemple, les mots *bom* « printemps », *gaeul* « automne », *salam* « homme », *jal* « bien », etc. sont des morphèmes autonomes, alors que les formes *meok-* (radical verbal), *aleum-* (radical adjectival), *-da* (suffixe terminal), *-dap* (suffixe de dérivation adjectivale), etc. sont des morphèmes non autonomes. Les morphèmes non autonomes comme *meok-*, *aleum-*, *-da*, etc. peuvent constituer des mots par combinaison avec d'autres morphèmes : *meok-da* « manger », *aleum-dap-da* « être beau », etc.

Nous présentons la définition des types d'éléments de mots. Les positions des linguistes sur les affixes sont assez consensuelles. En général, on admet que les affixes ne jouent pas un rôle de base dans la formation des mots et que ce sont des morphèmes non autonomes. Les préfixes n'ajoutent que des éléments de sens aux mots ou aux radicaux et ils n'ont pas de fonction de dérivation syntaxique ni de fonction de flexion. Les suffixes, quant à eux, se subdivisent en trois catégories d'après leur fonction : les

¹ *Morphème autonome* est traduit par *free morpheme* et *morphème non autonome* par *bound morpheme* en anglais.

suffixes de dérivation lexicale qui ajoutent des éléments de sens aux mots, les suffixes de dérivation syntaxique qui modifient la catégorie grammaticale et les suffixes de flexion qui jouent plusieurs rôles selon les catégories.

En revanche, selon les positions des linguistes coréens sur les racines, la terminologie utilisée est différente : racine, radical, tronc, noyau, base, etc., et il arrive que leurs frontières se superposent². Les racines, qui sont sémantiquement pleines et sont des mots simples ou des radicaux, sont à la base de la formation des mots. Nous distinguons les radicaux verbaux (ou adjectivaux) des radicaux sans catégorie grammaticale. Les radicaux verbaux *Rv* et les radicaux adjectivaux *Ra* se combinent directement avec les suffixes flexionnels. Par exemple, le radical verbal *meok-* « manger » se combine directement avec un suffixe flexionnel *-da*. En revanche, le radical *aleum-* « beau », qui demande obligatoirement un suffixe de dérivation adjectivale *-dap*, et le radical *moho-* « vague », qui demande un suffixe de dérivation, sont des radicaux sans catégorie grammaticale³. Nous appelons *Rx* ce type de radical qui n'a pas de catégorie grammaticale et qui n'entre dans la formation des mots que comme base. Par exemple :

<i>meok-da</i>	<i>Rv-St</i> « manger »
<i>aleum-dap-da</i>	<i>Rx-Sfa-St</i> « être beau »
<i>moho-ha-da</i>	<i>Rx-Sfa-St</i> « être vague »
<i>moho-seong</i>	<i>Rx-Sfn</i> « caractère vague »

Tous les radicaux *Rv*, *Ra*, et *Rx* ne sont pas des mots, tandis que les noms sont des mots. Ce point différent permet de distinguer les suffixes flexionnels⁴ des noms des suffixes verbaux ou adjectivaux : ceux-ci se combinent avec des radicaux non autonomes et ceux-là avec des mots autonomes. Les suffixes flexionnels des noms, à savoir les *postpositions nominales*⁵ sont une partie du discours dans la grammaire coréenne, tandis que les suffixes flexionnels du verbe et de l'adjectif ne le sont pas. Résumons les définitions des éléments qui interviennent dans la formation des mots par le tableau ci-dessous :

² Nous traduisons le mot coréen *eogeun* par radical, *eogan* par tronc, *eogi* par base. cf. Heo Ung (1983, 1995), Lee I.S (1968, 1975), Park E.J (1996), No M.H (1998), Nam G.S et Ko Y.K (1985), Jeong W.S (1992), Seo J.S (1995), etc.

³ Certains linguistes considèrent la forme *aleum-dap* comme un radical adjectival.

⁴ On les appelle souvent les suffixes syntaxiques.

⁵ Nous traduisons le mot coréen *josa* par postposition nominale.

		Morphème autonome	Combinaison avec les suffixes flexionnels
Base	Mot simple	Oui (Mot)	Directe
	<i>Rv / Ra</i>	Non (élément de mot)	Directe
	<i>Rx</i>	Non (élément de mot)	Non directe
Affixe	<i>Pf</i> ou <i>Sf</i> de dérivation	Non (élément de mot)	-
	<i>Sf</i> de flexion	Non (élément de mot)	-

Figure 2. La définition des éléments de formation des mots

Maintenant, considérons la formation des noms qui fait l'objet de notre travail. Les noms sont classés en noms simples et noms complexes. Le nom simple se définit comme toute unité autonome et non décomposable (par exemple, *eomeoni* « mère », *nai* « âge » ou *te* « anneau »). Tout nom qui n'est pas un nom simple est appelé nom complexe et comporte au moins deux éléments tels qu'un affixe, un nom simple ou un radical. Parmi les noms complexes, on distingue les noms dérivés et les noms composés.

Un nom dérivé comme *si-eomeoni* « belle-mère » est un nom qui constitue une unité lexicale nouvelle à partir d'une base et d'un affixe. Les préfixes et les suffixes de dérivation lexicale n'ajoutent qu'un élément de sens au substantif et les suffixes de dérivation syntaxique comme *-eum*, *-gi*, *-gae*, etc. modifient la catégorie grammaticale du substantif. Les suffixes flexionnels qui sont appelés les *postpositions nominales* dans notre travail ont plusieurs rôles selon leur catégorie : marquer les rôles syntaxiques des substantifs dans une phrase, lier des substantifs ayant la même fonction syntaxique dans la phrase ou ajouter une nuance sémantique à la phrase⁶. Les suffixes de dérivation syntaxique et les postpositions nominales sont très productifs. Un nom composé comme *nai-te* « anneau d'accroissement » est un nom qui constitue une unité sémantique mais qui est constitué de deux bases. Nous résumons ainsi la formation des noms :

- a. Noms simples : noms sans dérivation et sans composition
- b. Noms dérivés : base + affixe(s) de dérivation
- c. Noms composés :
 - (i) base + base
 - (ii) base + nom dérivé

⁶ cf. chapitre 4.

- (iii) nom dérivé + base
- (iv) nom dérivé + nom dérivé
- (v) nom composé + affixe

1-2. Exemples proprement coréens

Tout d'abord, donnons des exemples de noms dérivés (1). Les études sur les mots dérivés se font généralement sur le vocabulaire proprement coréen et quelques affixes sino-coréens, comme *si-* en (1-i) et *-ja* en (1-ii), qui sont intégrés dans le système coréen par leur comportement qui ressemble à celui des affixes proprement coréens.

(1) *Pf-N*, *N-Sf*, *Adv-Sf*, *Dét-Sf*, *V-sfn* ou *A-sfn*, *Rx-Sf*

(i) *Pf-N*

<i>si-eomeoni</i>	<i>Pf-N</i> « côté du mari » « mère » = « belle-mère »
<i>hol-si-eomeoni</i>	<i>Pf-Pf-N</i> « seul » « côté du mari » « mère » = « belle-mère qui est veuve »
<i>nal-gamja</i>	<i>Pf-N</i> « cru » « pomme de terre » = « pomme de terre crue »

(ii) *N-Sf*

<i>dalimi-jil</i>	<i>N-Sf</i> « fer à repasser » « action » = « repassage »
<i>bueok-ttegi</i>	<i>N-Sf</i> « cuisine » « suffixe péjoratif pour une personne » = « cuisinière » (expression vulgaire)
<i>deungdae-jigi</i>	<i>N-Sf</i> « phare » « gardien » = « gardien de phare »

(iii) *Adv-Sf*

<i>galo-daji</i>	« en largeur » « fermeture » = « porte ou fenêtre à glissière »
<i>magu-daji</i>	« imprudemment » « fermeture » = « action imprudente ou irréfléchie »
<i>magu-jabi</i>	« imprudemment » « action de saisir » = « action imprudente ou irréfléchie »

(iv) *Dét-Sf*

<i>i-li</i>	« ici » « direction » = « par ici »
<i>geu-li</i>	« là » « direction » = « par-là »
<i>jeo-li</i>	« là-bas » « direction » = « là-bas »

(v) **V-sfn** ou **A-sfn**

<i>jiu-gae</i>	V-sfn	« effacer »-sfn = « gomme »
<i>eol-eum</i>	V-sfn	« geler »-sfn = « glaçon »
<i>nop-i</i>	A-sfn	« être haut »-sfn = « hauteur »
<i>keu-gi</i>	A-sfn	« être grand »-sfn = « grandeur »
<i>gippeu-m</i>	A-sfn	« être joyeux »-sfn = « joie »

(vi) **Rx-Sf** ⁷

- a. *meongteong-guli* « sot » « (péjoratif) personne » = « sot, idiot »
malla-kkaengi « être maigre » « (suffixe vulgaire) personne » = « personne très maigre »
saechim-ttegi « feindre l'innocence » « (péjoratif) personne »
= « personne qui se donne des apparences réservées »
gaegu-jangi « être coquin » « (péjoratif) personne » = « petit coquin »
- b. *gwittul-ami* « cri du grillon »-sfn = « grillon »
gaegul-i « cri de la grenouille »-sfn = « grenouille »
kkobulang-i « apparence courbée »-sfn = « objet voûté ou courbé »
yaong-i « cri du chat »-sfn = « chat »

Il existe des noms dérivés qui comportent plusieurs affixes.

⁷ b'. *gwittul-gwittul* (**Rx-Rx = Adv**) « cri du grillon (onomatopée) »
gaegul-gaegul (**Rx-Rx = Adv**) « cri de la grenouille (onomatopée) »
kkobulang-kkobulang (**Rx-Rx = Adv**) « apparence courbée (onomatopée) »
yaong-yaong (**Rx-Rx = Adv**) « cri du chat (onomatopée) »

Les exemples (vi-b') sont des onomatopées. Certains linguistes distinguent les *Rx* en (vi-a) des radicaux adverbiaux en (vi-b) qui sont éléments d'onomatopées doubles (vi-b'). Mais nous regroupons sous l'abréviation *Rx* les exemples en (vi-a) et ceux en (vi-b) parce que les éléments d'onomatopées doubles peuvent constituer non seulement des adverbes ((vi-b')) mais aussi des noms avec les suffixes de nominalisation ((vi-b)) comme les autres *Rx*.

(vii) *mat-dambae-jil* **Pf-N-Sf** « en face de » « cigarette » « action » = « action de fumer en présence d'un adulte supérieur »

Dans les études traditionnelles sur le vocabulaire proprement coréen, les linguistes ont établi des critères de définition des noms composés⁸ : la modification du sens (*bam* "nuit" - *sonnim* "client" = *bamsonnim* "voleur"), les modifications phonologiques (*su* "mâle" - *gae* "chien" = *sukae* "chien masculin" ou *mal* "cheval" - *so* "bœuf" = *maso* "bétail"), l'insertion de la consonne spécifique « s »⁹ (*cho* "bougie" - *s* - *bul* "lumière" = *cho-s-bul* "flamme"), les restrictions sur l'ordre des éléments (*bam* "nuit" - *nat* "jour" = *bamnat* "toujours" ; **natbam*), etc. Nous n'entrons pas dans le détail des critères de définition des noms composés.

Quant aux conventions typographiques sur les noms composés proprement coréens, la soudure typographique peut jouer un rôle de marqueur de composition en coréen comme le trait d'union en français ou anglais. Nous présentons ci-dessous des exemples de noms composés proprement coréens.

(2) *N-N, Dét-N, Asfd-N, Vsfd-N, Adv-Adv, N-s-N, Adv-N, Ra-N, Rv-N et Rx-N*

(i) *N-N*

<i>kkot-galu</i>	<i>N-N</i>	« fleur » « poudre » = « pollen »
<i>dukkeobi-jip</i>	<i>N-N</i>	« crapaud » « maison » = « boîte à fusibles »
<i>neolbi-ttwigi</i>	<i>Asfn-Vsfn</i>	« largeur » « saut » = « saut en longueur »
<i>nopi-ttwigi</i>	<i>Asfn-Vsfn</i>	« hauteur » « saut » = « saut en hauteur »
<i>ttwim-ttwigi</i>	<i>Vsfn-Vsfn</i>	« course » « saut » = « course »
<i>maejeum-mal</i>	<i>Vsfn-N</i>	« terminaison » « parole » = « conclusion »
<i>swim-pyo</i>	<i>Vsfn-N</i>	« pause » « signe » = « pause »
<i>bap-beoli</i>	<i>N-Vsfn</i>	« riz » « gain » = « gagne-pain »
<i>nun-useum</i>	<i>N-Vsfn</i>	« yeux » « rire » = « regard souriant »

(ii) *Dét-N*

sae-hae « nouveau » « année » = « nouvelle année »

⁸ cf. Voir les détails Seo J.S (1995), Park E.J (1996) et Kim G.H (1994).

⁹ Cette consonne « s ; *siot* » est appelée *sai siot* (consonne intermédiaire). Voir les exemples de (2-vi) pour les détails.

sae-eonni « nouveau » « sœur » = « belle-sœur »
cheot-salang « premier » « amour » = « premier amour »
i-geot « ce » « chose » = « ceci »

(iii) **Asfd-N**

keun-hyeung « être grand »-sfd « frère » = « frère aîné »
jakeun-abeoji « être petit »-sfd « père » = « oncle, frère cadet du père »
jakeun-jip « être petit »-sfd « maison » = « maison du frère cadet »

(iv) **Vsfd-N**

jukeul-him « mourir »-sfd « force » = « efforts à la mort »
tteun-guleum « flotter »-sfd « nuage » = « quelque chose d'éphémère comme un nuage flottant dans le ciel »
meokeul-geoli « manger »-sfd « matière » = « aliment, provisions »

(v) **Adv-Adv**

jal-mot « bien » « ne...pas ; mal » = « faute »

Grammaticalement et morphologiquement, les exemples de (i) - (v) obéissent aux règles ordinaires de combinaison pour les groupes nominaux du coréen. Dans la tradition linguistique du coréen, on appelle *composition syntaxique* la combinaison normale entre deux formes autonomes. Parmi les types de noms composés (i) - (v), le premier type *NN* est le plus nombreux et productif, alors que le dernier type *Adv-Adv* est le moins nombreux.

Alors que les types (i) - (v) relèvent de la composition syntaxique, les types (vi) - (xi) relèvent de la *composition déviante*. C'est-à-dire qu'ils n'obéissent pas aux règles grammaticales et morphologiques de combinaison et qu'ils n'apparaissent que dans des noms composés, et non dans des groupes nominaux libres.

(vi) **N-s-N**

ko-s-mul « nez » « eau » = « morve »
cho-s-bul « bougie » « lumière » = « flamme »

chima-s-balam « jupe » « vent » = « ingérence des mères d'élèves dans l'administration des écoles primaires »
bada-s-ga « mer » « bord » = « plage »

Dans les exemples (vi), la consonne « s », dite consonne intermédiaire, est insérée entre les deux mots. Ce « s » est appelé un marqueur de nom composé parce qu'il n'apparaît pas dans les groupes nominaux libres¹⁰.

Les exemples de (vii) sont des compositions déviantes parce que les *Ra* ou les *Rv* se combinent directement avec un nom sans suffixe déterminatif. Dans la composition syntaxique, lorsqu'un verbe ou un adjectif qualifie un nom, il est ordinairement suivi d'un suffixe déterminatif (cf. les exemples de (iii)-(iv)).

(vii) ***Ra-N*** ou ***Rv-N***

<i>neut-gaeul</i>	<i>Ra-N</i>	« tardif, dernier » « automne » = « fin de l'automne »
<i>geom-beoseot</i>	<i>Ra-N</i>	« noir » « champignon » = « taches noires sur la peau des vieillards »
<i>jeop-kal</i>	<i>Rv-N</i>	« plier » « couteau » = « canif »
<i>mat-balam</i>	<i>Rv-N</i>	« recevoir » « vent » = « vent contraire ou de face »

Dans les exemples suivants, des éléments d'onomatopées *Rx* se combinent avec des noms.

(viii) ***Rx-N***

<i>sandeul-balam</i>	« manière légère » « vent » = « vent léger »
<i>heoltteok-gogae</i>	« manière de haleter » « col » = « col très escarpé à monter »
<i>munggae-guleum</i>	« apparence en couches épaisses » « nuage » = « nuage en couches épaisses »
<i>jongjong-geoleum</i>	« manière de marcher à petits pas rapides » « pas » = « petits pas rapides »
<i>heundeul-uija</i>	« apparence de se balancer continuellement » « chaise » = « fauteuil à bascule »

¹⁰ Lee I.S et Nam G.S (1987), p. 46.

Le premier élément de chacun de ces exemples n'est pas un adverbe et son radical n'est pas autonome. Pour obtenir un mot autonome autonomie, il faut le redoubler : *sandeul-sandeul*, *heoltteok-heoltteok*, etc.

Les exemples suivants comportent un adverbe et un nom. Cette combinaison est peu naturelle en coréen sauf entre un adverbe intensif et un nom intensif (e.i., *aju buja Adv N* "un très riche").

(xi) **Adv-N**

nalanhi-kkol « parallèlement » « forme » = « parallélogramme »

kkalkkal-useum « son d'éclat de rire » « rire » = « éclat de rire »

1-3. La formation des mots sino-coréens

Dans le vocabulaire sino-coréen, la formation des mots présente des différences par rapport aux mots proprement coréens. La plupart des études sur les mots composés ont porté sur le vocabulaire proprement coréen et les mots composés sino-coréens ont été peu étudiés de façon systématique dans la grammaire coréenne¹¹.

Chaque idéogramme chinois emprunté est associé à une forme monosyllabique et à une signification, mais en général il n'est pas autonome dans le système du coréen sauf dans le cas exceptionnel de noms sino-coréens comme *mun* « porte » ou *chaek* « livre » qui sont autonomes et se comportent syntaxiquement comme des noms dans la phrase puisqu'il n'existe pas de mots proprement coréens correspondants.

(3) *mun*-(*i* + *eul* + *eulo* + ...) = « porte »-(*Nmtf* + *Acc* + *Postp.eulo* + ...)

chaek-(*i* + *eul* + *eulo* + ...) = « livre »-(*Nmtf* + *Acc* + *Postp.eulo* + ...)

En revanche, dans les exemples de (4), *wi* « grand » et *in* « homme », qui sont issus de mots dans le système du chinois, deviennent des formes non autonomes en coréen, même si elles gardent leur sens plein :

(4) a. *wi-in* (**Rx-Rx** = *N*) « grand » « homme » = « grand homme »

¹¹ Ils ont souvent été étudiés dans la grammaire chinoise à l'extérieur du système du coréen. Etant donné que le vocabulaire sino-coréen est historiquement emprunté au chinois, il combine des propriétés du chinois et d'autres qu'il a acquises en s'assimilant au coréen.

- b. **wi*-(*ga* + *da*) (*Rx*-(*Nmtf* + *St*)) « grand »-(*Nmtf* + *St*)¹²
 **in*- (*Rx*-*Nmtf*) « homme »-*Nmtf*
 c. *wi-in-i* ((*Rx-Rx*)-*Nmtf* = *N*-*Nmtf*) « grand homme » -*Nmtf*

La plupart des noms sino-coréens sont formés par une combinaison de plusieurs idéogrammes, en général deux. Au niveau sémantique, la plupart des noms dissyllabiques sino-coréens sont traditionnellement classés parmi les composés.

Les idéogrammes empruntés au chinois peuvent devenir des radicaux non autonomes ou des noms monosyllabiques (3). Ils peuvent aussi devenir des affixes monosyllabiques : -*in* « homme » est un radical dans (5a), mais un suffixe dans (5b).

- (5) a. *in-gan* (*Rx-Rx* = *N*) « homme » « entre » = « homme »
 wi-in (*Rx-Rx* = *N*) « être grand » « homme » = « grand homme »
 b. *ye-sul* (*Rx-Rx* = *N*) « arts » « science » = « art »
 ye-sul-in ((*Rx-Rx* = *N*)-*Sf*) « arts » « science » « homme » = « artiste »

Si on examine les exemples suivants, on perçoit la difficulté de distinguer un mot monosyllabique, un radical monosyllabique et un affixe monosyllabique.

- (6) a. *nae-bunbi* (*Pf* ou *Rx*)-*N* « intérieur » « sécrétion » = « sécrétion interne »
 nae-chulhyeol (*Pf* ou *Rx*)-*N* « intérieur » « hémorragie » = « hémorragie interne »
 b. *gu-nae* (*Rx-Rx* = *N*) « quartier » « intérieur » = « enceinte »
 gyo-nae (*Rx-Rx* = *N*) « école » « intérieur » = « intérieur d'une école »
 c. *guyeok nae* *N* (*Ni* ou -*Sf*)¹³ « zone » « intérieur » = « intérieur d'une zone »
 jiban nae *N* (*Ni* ou -*Sf*) « famille » « intérieur » = « intérieur de la famille »

La forme *nae* est souvent traitée comme un préfixe dans (6a), un radical dans (6b) et un nom incomplet dans (6c). Le sens est le même. En outre, on n'a pas de critère précis pour distinguer ces trois emplois : (6a) pourrait aussi être qualifié de radical et (6c) de suffixe¹⁴.

¹² Si la forme *wi* était un nom en coréen, elle pourrait se combiner avec la postposition du nominatif. Si elle était un adjectif en coréen, elle pourrait se combiner avec le suffixe terminal comme les autres adjectifs.

¹³ Si l'élément *nae* était un suffixe, on ne mettrait pas d'espace.

¹⁴ Cette distinction est très importante pour la détermination de l'espacement. En coréen, si une séquence composée d'un déterminant et d'un nom n'est pas un nom composé au sens traditionnel, on met

On se heurte à une difficulté de délimitation entre radical, préfixe et déterminant en position de premier élément du mot, et entre radical, suffixe et nom incomplet en position de dernier élément du mot. No M.H (1998) propose une série de critères pour distinguer les différents emplois. Les critères de distinction entre déterminant et autres éléments (préfixe ou radical) sont la possibilité d'insertion d'autres éléments, les restrictions distributionnelles, la portée de la qualification, l'impossibilité de substitution, etc. Les critères de distinction du préfixe par rapport au radical sont la modification du sens original, la productivité, la compatibilité avec les mots proprement coréens, les modifications syntaxiques de la base, etc. Les critères mentionnés fonctionnent bien sur les exemples de (7) :

- (7) a. *oe-samchon* *Pf-N* « côté de la femme » « oncle » = « frère de la mère »
 si-eomeoni *Pf-N* « côté du mari » « mère » = « belle-mère »
 b. *oe-ga* *Rx-N* « côté de la femme » « famille » = « famille de la mère »
 si-daek *Rx-N* « côté du mari » « famille du mari » = « famille du mari »
 c. *bon sageon* *Dét N* « principal ; ce » « affaire » = « cette affaire »
 gwi hoesa *Dét N* « votre ; vis-à-vis de l'interlocuteur avec politesse »
 « entreprise » = « votre entreprise »

Mais il n'est pas difficile de trouver des cas où l'on ne peut pas distinguer les trois emplois avec précision à l'aide de ces critères. Il s'agit de cas limite où la différence entre les trois emplois n'est pas claire :

- (8) a. *bon-buin* (*Pf* ou *Rx*)-*N* « principal » « femme » = « épouse officielle »
 gwi-geumsok (*Pf* ou *Rx*)-*N* « précieux, noble » « métal » = « métaux précieux »
 b. *chong-suip* (*Rx* ou *Dét*)-*N* « total » « revenu » = « revenu brut »
 gu-sidae (*Rx* ou *Dét*)-*N* « vieux » « période » = « ancien temps »
 c. *chong saengsanlyang* (*Dét* ou *Pf*)-*N*¹⁵ « total » « quantité de production » =
 « production totale »
 gu cheongsa (*Dét* ou *Pf*)-*N* « vieux » « bâtiment de l'administration » = « vieux
 bâtiment de l'administration »

Un radical sino-coréen monosyllabique en position de premier élément du mot a une

obligatoirement un blanc entre eux. Et on met obligatoirement un blanc entre un nom incomplet et son complément.

¹⁵ Si les éléments *chong* ou *gu* étaient des préfixes, on ne mettrait pas de blanc.

tendance diachronique à être perçu comme un préfixe ou comme un déterminant.

La situation est la même en position de dernier élément du mot. Considérons *-ja* « personne » en combinaison avec divers éléments.

- (9) a. *gaip-ja* *N-Sf* « abonnement » « personne » = « membre abonné »
gisul-ja *N-Sf* « technique » « personne » = « technicien »
 b. *bujae-ja* *Ra-Sf* « absence » « personne » = « absent »
yeongse-ja *Ra-Sf* « petits biens » « personne » = « personne pauvre »
 c. *bogyun-ja* *Rx-Sf* « porteur de bacilles » « personne » = « porteur de bacilles »
hugye-ja *Rx-Sf* « succession » « personne » = « successeur »

No M.H (1998) propose une série de critères pour distinguer les radicaux des suffixes et des noms incomplets en position de dernier élément du mot : le degré de dépendance de la base, les modifications syntaxiques de la base, la modification du sens, l'impossibilité d'occurrence en position de premier élément du mot, la compatibilité avec une unité plus grande que le mot, les restrictions sur la compatibilité de quelques postpositions nominales, la productivité, l'ordre des éléments, etc. Les exemples de (10) illustrent raisonnablement bien ces critères.

- (10) a. *yesul-ga* *N-Sf* « art » « personne » = « artiste »
gisul-ja *N-Sf* « technique » « personne » = « technicien »
 b. *aeguk-sim* *N-Rx* « patriotisme » « cœur » = « sentiment patriotique »
oeguk-eo *N-Rx* « étranger » « langue » = « langue étrangère »
 c. *hoeui jung* *N Ni* « séance » « milieu » = « (au) milieu de la séance »
naenyeon cho *N Ni* « an prochain » « début » = « début de l'an prochain »

Cependant, il existe aussi des cas moins clairs :

- (11) a. *hakgi mal* *N-(Rx ou Ni)*¹⁶ « semestre » « fin » = « fin de semestre »
yesang oe *N-(Rx ou Ni)* « prévision » « en dehors de » = « imprévision »
 b. *byeong-ja* *N-(Rx ou Sf)* « maladie » « personne » = « malade »
bu-ja *Rx-(Rx ou Sf)* « richesse » « personne » = « riche »
 c. *gujo sang* *N (Ni ou -Sf)*¹⁷ « structure » « point » = « (sur le) point structural »
hyeongsik sang *N (Ni ou -Sf)* « forme » « point » = « (sur le) point formel »

¹⁶ Si les éléments *mal* et *oe* étaient des *Rx*, on ne mettrait pas de blanc.

¹⁷ Si *sang* était un suffixe, on ne mettrait pas de blanc.

Le statut des idéogrammes, mis à part les noms monosyllabiques comme (3), est souvent vague malgré une série de critères. Ce problème n'est pas clairement résolu dans la linguistique coréenne traditionnelle.

2. Les noms composés du point de vue du traitement informatique

2-1. Problèmes spécifiques du traitement des noms composés

Nous allons voir comment les noms composés sont traités dans la tradition informatique. Il semble que le terme de nom composé ne soit pas utilisé pour désigner les mêmes formes en informatique et en linguistique traditionnelle. Le terme « nom composé » est tiré de la tradition linguistique, mais le « traitement des noms composés » en l'informatique vise en fait les séquences constituées de plusieurs noms (*NN*, *NNN*, *NNNN*,...) : ce qui couvre à la fois les noms composés et les groupes nominaux libres, sans distinction. Pourquoi cette situation? Nous allons examiner les quatre problèmes spécifiques des séquences du type *NN*⁺ du point de vue des traitements informatiques.

Tout d'abord, l'agglutination entre les noms (surtout sino-coréens) est très libre. Les divers exemples illustrent la différence entre noms composés et groupe nominal libre :

- (12) a. [*hucheonseong myeonyeok gyeolpipjeung*] *yebang* (*NNNN*)
 « qualité acquise » « immunité » « syndrome de déficience » « prévention »
 = « prévention du *SIDA* (Syndrome Immuno-Déficitaire Acquis) »
 b. [*gukhoe uijang*] *seongeo gaesi* (*NNNN*)
 « l'Assemblée nationale » « président » « élection » « commencement »
 = « commencement de l'élection du président de l'Assemblée nationale »

Ces séquences sont des groupes nominaux libres constitués de quatre noms, mais toutes les deux comportent un nom composé : « *SIDA* » (trois noms), « président de l'Assemblée nationale » (deux noms). En coréen, les séquences composées de type *NN*⁺, c'est-à-dire *NN*, *NNN*, *NNNN*..., constituent la grande majorité des composés nominaux. Les combinaisons des noms sont plus libres en coréen qu'en français et sont en nombre théoriquement infini. En fonction de leur degré de figement, les séquences de noms forment un continuum entre les noms composés et les groupes nominaux libres. De plus, il n'y a pas de critères précis applicables informatiquement de distinction entre eux.

Deuxièmement, les séquences composées de type NN^+ sont très nombreuses. Comme nous l'avions mentionné dans la section 1-1, les noms composés de type NN sont les plus nombreux. D'après notre étude, le nombre de noms composés communs de type NN^+ à espacement facultatif comme en (12) est d'au moins 45000. Selon le rapport du *Projet Sejong 21* (1998), le nombre des noms propres composés de type NN^+ à espacement facultatif est de 30000. On estime que le nombre de noms composés techniques de type NN^+ à espacement facultatif est de plusieurs dizaines de milliers. Enfin, on ne peut pas même estimer le nombre des groupes nominaux libres de type NN^+ .

Troisièmement, les conventions typographiques d'espacement compliquent ce problème. Voici un exemple où l'espacement entre les noms est variable. Les deux exemples de (12) peuvent se réaliser graphiquement de huit façons chacun :

$$(12') \text{ a. } [\textit{hucheonseong} \quad \textit{myeonyek} \quad \textit{gyeolpipjeung}] \textit{yebang} \quad [N \quad N \quad N] \quad N$$

$$\qquad \qquad \qquad 2 \qquad \quad \times \qquad 2 \qquad \quad \times \qquad 2 \qquad \quad = \quad 8$$

$[NNN]N$	$[NNN] \quad N$
$[N \quad NN]N$	$[N \quad NN] \quad N$
$[N \quad N \quad N]N$	$[N \quad N \quad N] \quad N$
$[NN \quad N]N$	$[NN \quad N] \quad N$

후천성면역결핍증예방	후천성면역결핍증 예방
후천성 면역결핍증예방	후천성 면역결핍증 예방
후천성 면역 결핍증예방	후천성 면역 결핍증 예방
후천성면역 결핍증예방	후천성면역 결핍증 예방

$$\text{b. } [\textit{gukhoe} \quad \textit{uijang}] \quad \textit{seongeo} \quad \textit{gaesi}^{18}$$

$[N \quad N] \quad N \quad N$? $[N \quad N]N \quad N$
$[N \quad N] \quad NN$	$[N \quad N]NN$
$[NN] \quad N \quad N$	$[NN]N \quad N$
$[NN] \quad NN$	$[NN]NN$

국회 의장 선거 개시	? 국회 의장선거 개시
국회 의장 선거개시	국회 의장선거개시

¹⁸ La forme $[N \quad N]N \quad N$ est un peu douteuse.

국회의장 선거 개시	국회의장선거 개시
국회의장 선거개시	국회의장선거개시

Les noms composés *hucheonseong myeonyek gyeolpipjeung* et *gukhoe uijang* sont des mots sino-coréens. Tandis que pour les composés nominaux sino-coréens l'espacement entre les noms est souvent facultatif, pour les noms composés proprement coréens au sens traditionnel on ne met pas de blanc dans les dictionnaires éditoriaux, mais il arrive que l'on mette un blanc dans l'usage selon le nombre de syllabes des séquences (cf. voir le chapitre 2).

Quatrièmement, en informatique, le problème de la segmentation correcte des séquences nominales non enregistrées dans un dictionnaire est toujours ouvert. Si les groupes nominaux libres sont soudés, ils peuvent être absents du dictionnaire électronique même complet. Pour réduire le nombre des séquences nominales inconnues, il faut obligatoirement une procédure de segmentation. La segmentation correcte des séquences soudées du type de NN^+ du coréen est un processus indispensable en raison de diverses applications comme la traduction automatique, la correction de l'espacement, l'indexation automatique, l'extraction de mots-clés, etc.

Les deux premiers problèmes viennent de la productivité très élevée des noms et des groupes nominaux et les deux derniers problèmes de l'espacement irrégulier entre les mots. Il nous semble que la segmentation des séquences nominales permet de résoudre plus ou moins à la fois le problème de la productivité et celui de l'espacement irrégulier. En fait, l'histoire du traitement des noms composés du coréen en informatique est celle de la découverte de méthodologies de décomposition des séquences nominales. Nous allons examiner les travaux apparentés sur les noms composés coréens en informatique.

2-2. Travaux apparentés sur les noms composés

Le problème de la segmentation des séquences nominales se résume en trois points :

- (1) Génération des segmentations candidats pour chaque séquence nominale soudée
- (2) Choix de la meilleure segmentation en cas d'ambiguïté
- (3) Analyse des séquences comprenant des noms inconnus

En général, lors de l'analyse morphologique, après avoir reconnu les séquences

nominales (soit les noms simples soit les noms composés) qui sont répertoriées dans le dictionnaire électronique, les systèmes segmentent les séquences nominales soudées avec leur propre méthode pour leur propre application. La méthode par structure syllabique du coréen est souvent utilisée pour le découpage des séquences nominales soudées. Nous présentons ici les trois méthodes de décomposition : la décomposition algorithmique (Kang S.S, 1993 et 1998), la décomposition d'après les informations syllabiques dans un dictionnaire électronique (Choi J.H, 1996) et la décomposition par les statistiques (Yoon B.H et al., 1995).

Kang S.S (1993) a examiné la typologie syllabique de 18283 entrées de noms composés tirées des dictionnaires éditoriaux et comparé les fréquences des différents types : $2+2^{19}$ constituent 90% des noms composés à 4 syllabes (73.1% sur 18282 entrées), $2+3$ ou $3+2$ constituent 62% des noms composés à 5 syllabes (17.4 % sur 18282 entrées) et $2+2+2$ ou $3+3$ constituent 53% des noms composés à 6 syllabes (6.2% sur 18282 entrées). Il prévoit que si un système identifie les noms composés du type $2+2$, $2+3$, $3+2$, $2+2+2$ et $3+3$, il reconnaîtra 80% des séquences nominales soudées dans le corpus, car les noms composés à 4-6 syllabes constituent 96.7% des noms composés sur 18282 entrées.

Kang S.S (1998) présente deux règles pour éviter des segmentations inutiles.

(i) Deux règles d'évitement qui permettent de ne pas passer par un processus de décomposition

- Si une séquence nominale à 4-5 syllabes est enregistrée dans le dictionnaire, a priori, on la garde sans découpage intérieur. Par exemple, *chodeung-hakgyo* ($2+2$) « école élémentaire » qui est enregistré dans le dictionnaire n'est pas segmenté en $2+2$.
- Si une séquence nominale reste inconnue après la consultation du dictionnaire, on parcourt à nouveau une (ou plusieurs) syllabe(s) de cette séquence dans le dictionnaire en sautant syllabe par syllabe. Si l'on trouve une certaine partie de cette séquence dans le dictionnaire, on considère la partie sautée comme un affixe inconnu ou comme un terme inconnu.

Pour un nom composé à n syllabes, le nombre maximum de possibilités de segmentation est de 2^{n-1} en insérant le(s) blancs entre les syllabes mais en général une

¹⁹ Ici, le chiffre indique le nombre de syllabes.

seule de ces possibilités est bonne²⁰. Kang S.S (1998) présente quatre règles pour produire des segmentations plausibles.

(ii) Quatre règles de décomposition qui aident à réduire le nombre de 2^{n-1} segmentations candidats

- générer, à priori, les segmentations candidats en se référant au dictionnaire
- ajouter des segmentations candidats correspondant aux types fréquents 2+2, 2+3, 3+2, 2+2+2 et 3+3
- la segmentation ne doit pas être monosyllabique. Si l'on trouve un nom monosyllabique, alors il est considéré comme un affixe
- si une séquence nominale de longueur n commence par un nom à deux syllabes, ajouter une segmentation commençant par une séquence à trois syllabes parce qu'une séquence à trois syllabes est souvent constituée d'un nom à deux syllabes et d'un suffixe monosyllabique. Pour ce type de séquence, il y a donc au moins deux candidats : $2 + n-2$ et $3 + n-3$

En engendrant les segmentations candidats plausibles à l'aide des règles ci-dessus, le système donne à chaque candidat un poids, puis il sélectionne le meilleur candidat selon son poids.

Choi J.H (1996) a aussi utilisé la méthode par nombre de syllabes pour deux applications : l'indexation automatique et la correction d'espacement. Il enregistre le nombre de syllabes pour chaque entrée dans un dictionnaire électronique. Cette méthode permet de consulter le dictionnaire une seule fois lors de l'analyse morphologique et d'éviter une seconde consultation du dictionnaire pour la segmentation. Voici quelques entrées extraits du dictionnaire des séquences nominales²¹ :

ga-gam-seung-je
ga-gam-sok-jeon-dong-gi_4 => *ga-gam-sok jeon-dong-gi*
ga-gam-gye-do-pyo-si_35 => *ga-gam gye-do pyo-si*
ga-lo-bang-hyang-su-jeong-ja-seok_357
 => *ga-lo bang-hyang su-jeong ja-seok*

Dans ces exemples, le chiffre après « _ » indique la position des espacements. Par

²⁰ Par exemple, pour un nom composé à 4 syllabes, on peut générer $2^3 (= 8)$ segmentations candidats en insérant des blancs.

²¹ « => » indique la segmentation enregistrée dans le dictionnaire électronique.

exemple, « _4 » indique un espacement avant la quatrième syllabe. Si l'on n'a pas réussi à faire l'analyse morphologique par dictionnaire, les séquences nominales inconnues entrent dans un processus de segmentation qui dépend du nombre de syllabes. Ici, nous présentons l'algorithme prévu pour les séquences nominales à quatre syllabes :

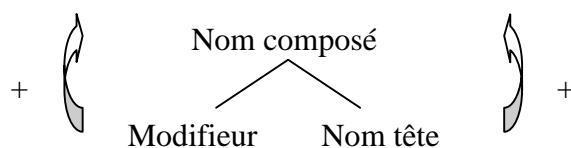
```
int divide_compound_nouns_4(unsigned word1[])
/* Arrays for
   F[1]: front one syllable, F[2]: front two syllables, F[3]: front three syllables,
   B[1]: rear one syllable, B[2]: rear two syllables, B[3]: rear three syllables */
{
  if (success_search_dic(F[4])
      segment according to the spacing information;
  else if (success_search_dic(F[2])&&success_search_dic(B[2]))
      if (B[2] == juui, namu, jali)22 segment 4_0;
      else segment 2_2;
  else if (success_search_dic(F[3])) {
      if (B[1] == suffix) segment 4_0;
      else if (success_search_dic(B[1])) segment 3_1;
  }
  else if (success_search_dic(B[3])&&F[1] == prefix) segment 4_0;
  else segment 4_0; /*default*/
}
```

Pour lever l'ambiguïté de la segmentation, Yoon B.H et al. (1995) présentent une méthode de décomposition par les statistiques utilisant la fréquence du nom tête dans les séquences nominales ou la fréquence des structures qui spécifient le nombre d'éléments et le nombre de syllabes de chaque élément. Le système génère des candidats en se référant au dictionnaire. Si le nombre du candidat est un, il sort. Sinon, le système sélectionne le meilleur candidat en se référant à la fréquence du nom tête ou à la fréquence des structures par nombre syllabique.

Regardons le cas où la fréquence du nom tête dans les séquences nominales est utilisée pour lever l'ambiguïté de la segmentation. Le nom composé est constitué d'un modifieur et d'un nom tête, et le modifieur ou le nom tête peuvent récursivement être

²² Il nous semble qu'il a considéré comme noms dérivés ou noms composés au sens traditionnel les séquences nominales se terminant par *juui, namu, jali* qui signifient « -sme », « arbre » et « constellation (dans l'astrologie) ». Il ne veut donc pas les segmenter.

constitué d'un sous-modifieur et d'un sous-nom tête :



Il a pris 81276 entrées d'un dictionnaire éditorial²³ où chaque entrée est constituée de $N_1 N_2 \dots N_{n-1} N_n$, puis il a acquis l'information statistique de la façon suivante :

- (i) Le premier nom N_1 est considéré comme un modifieur.
- (ii) Les noms N_2, \dots, N_{n-1} peuvent être considérés comme modifieurs ou noms têtes.
- (iii) Le dernier nom N_n est considéré comme un nom tête.

Par exemple, si une séquence nominale à 5 syllabes a deux segmentations ambiguës (2+3 ou 3+2) en se référant au dictionnaire d'après les structures syllabiques, le deuxième nom *bujoli* (13'a) ou *joli* (13'b) devient le nom tête de la séquence.

- (13) *sa-hoe-bu-jo-li* « absurdité de la société »
 (13') a. *sa-hoe bu-jo-li* « société » « absurdité » = « absurdité de la société »
 b. *sa-hoe-bu jo-li* « section ou département social » « cuisine » = « ? »

Or, *bujoli* de (13) dans les séquences nominales est utilisé 12 fois comme nom tête dans le dictionnaire éditorial, tandis que *joli* est utilisé 5 fois comme modifieur et une fois comme nom tête. En utilisant ces informations statistiques, on peut correctement segmenter la séquence (13) suivant (13'a).

Regardons un autre cas où la fréquence des structures par nombre syllabique est utilisée. Par exemple, la séquence (14) à 6 syllabes a deux segmentations ambiguës en se référant au dictionnaire (2+2+2 ou 3+3)²⁴ :

- (14) *dae-hak-saeng-seon-gyo-hoe* « étudiant » « réunion de la mission » = « réunion de la mission des étudiants »

²³ Il a utilisé *Keumsungpan Gukeo Daesajeon* (Grand dictionnaire du coréen dans la version *Keum-sung Publ.Co.*)(Kim S. H et al., 1995), Keum-sung Publ.Co.

²⁴ Pour la séquence *dae-hak-saeng-seon-gyo-hoe*, il n'existe pas les segmentations 4+2 ou 2+4 en se référant au dictionnaire.

- (14') a. *dae-hak saeng-seon gyo-hoe* « université » « poisson » « église » = « ? »
 b. *dae-hak-saeng seon-gyo-hoe* « étudiant » « réunion de la mission » =
 « réunion de la mission des étudiants »

Selon son expérience, il y avait 37 segmentations ambiguës pour les séquences à 6 syllabes qui sont ambiguës entre 2+2+2 et 3+3, et la segmentation « 3+3 » était toujours correcte pour les 37 cas. Il applique cette information pour le cas de segmentations ambiguës (2+2+2 ou 3+3) des séquences nominales à 6 syllabes et on peut donc choisir la structure 3+3 (14'b) comme résultat correct de la segmentation suivant la fréquence des structures par nombre syllabique. Selon les cas, Yoon B.H alterne les deux informations mentionnées (la fréquence du nom tête ou la fréquence des structures par nombre syllabique) pour lever l'ambiguïté de la segmentation.

Il existe d'autres méthodes de segmentation des noms composés sans utiliser l'information syllabique. Jang D.H et Maeng S.H (1996) utilise l'apprentissage automatique pour l'indexation effective. Le système apprend le lexique du corpus, et puis il utilise le résultat de cet apprentissage lors du découpage des séquences nominales. Cette méthode permet de réduire le travail à la main pour la construction du dictionnaire et ne demande pas l'usage d'un analyseur morphologique.

Yoon J.T et al (1998) aussi utilisent le corpus entraîné par apprentissage. Mais ils essaient d'analyser la structure des séquences nominales en utilisant des informations sur la relation entre les noms : les deux noms sont liés par la relation prédicat/argument ou modifieur/nom tête. Ces deux relations sont extraites du *Yonsei*, corpus étiqueté qui contient 30 millions d'unités typographiques.

2-3. Discussion

Les chercheurs issus du monde de l'informatique montrent rarement de l'intérêt pour les dictionnaires eux-mêmes. Dans la tradition informatique, rappelons que lorsqu'une séquence est composée de plus de deux noms (soit libre soit figé), c'est un nom composé. L'étude des noms composés est plutôt focalisée sur la décomposition automatique de formes en séquences de noms. On s'intéresse donc à trouver le découpage des séquences nominales à l'aide de dictionnaires éditoriaux ou par apprentissage automatique à partir de gros corpus. Ces études sur la décomposition sont très utiles pour reconnaître les séquences nominales soudées inconnues.

Or, ces travaux utilisent toutes les entrées des dictionnaires éditoriaux sans considération de leur figement linguistique. Maintenant, il y a des essais d'extraire

automatiquement les séquences nominales de corpus selon des relations sémantiques. Lee K.S et al. (2000) présentent les relations sémantiques entre deux noms : les relations entre le sujet et l'objet, entre l'objet et le prédicat, entre la totalité et la partie, entre la matière et l'objet, la relation possessive, la position, la causalité, etc. Par exemple, *beoseu sago (NN)* (bus - accident = accident par bus) est pris comme exemple de relation de causalité. Pourtant, si on ne distingue pas les noms composés des groupes nominaux libres, il est pratiquement impossible de passer en revue toutes les possibilités de séquences nominales du type de NN^+ . À notre avis, il est indispensable de distinguer les séquences figées des séquences libres et de lister les séquences figées, puisqu'il est pratiquement impossible de dresser la liste de toutes les séquences nominales dans les dictionnaires à cause de leur énorme productivité.

Quant aux variantes orthographiques ou liées à l'espacement dans l'usage, elles ne sont jamais prises en compte sérieusement. Les systèmes s'en tiennent aux conventions d'un ou plusieurs dictionnaires éditoriaux. Il n'y a pas d'études systématiques sur les variations d'espacement des séquences nominales, ni dans l'usage, ni dans les dictionnaires éditoriaux. Pourtant, des informations précises sur les conventions typographiques possibles en coréen sont indispensables. Nous allons voir les désaccords entre les dictionnaires éditoriaux et l'usage ou même entre les dictionnaires éditoriaux eux-mêmes dans le chapitre 2.

3. Deux dictionnaires électroniques des noms fondés sur la formation des mots

Comme nous l'avons dit, l'étude des noms dans la tradition linguistique coréenne est focalisée sur la formation des noms et par conséquent, les projets de construction de dictionnaires électroniques des noms sont naturellement focalisés sur les noms simples, les noms dérivés et les noms composés²⁵. Dans le prolongement de la tradition linguistique, il est intéressant de codifier toutes les informations étymologiques sur la composition des mots. Nous présentons deux perspectives représentatives sur le dictionnaire électronique des substantifs du coréen, qui sont fondamentalement basées sur la formation des mots : celle du *Projet Sejong 21* (1998, 1999, 2000, 2001) et celle de Nam Jee-Sun (1994, 1997).

Le dictionnaire électronique des noms dans le *Projet Sejong 21*, qui est un projet pour la construction d'un lexique du coréen, a tendance à respecter fidèlement le point de vue de la linguistique coréenne sur la formation des mots. Voici la structure du

²⁵ cf . section 1-1, pp. 8-9.

dictionnaire électronique des substantifs (Rapport du *Projet Sejong 21* (1998), p. 3) :

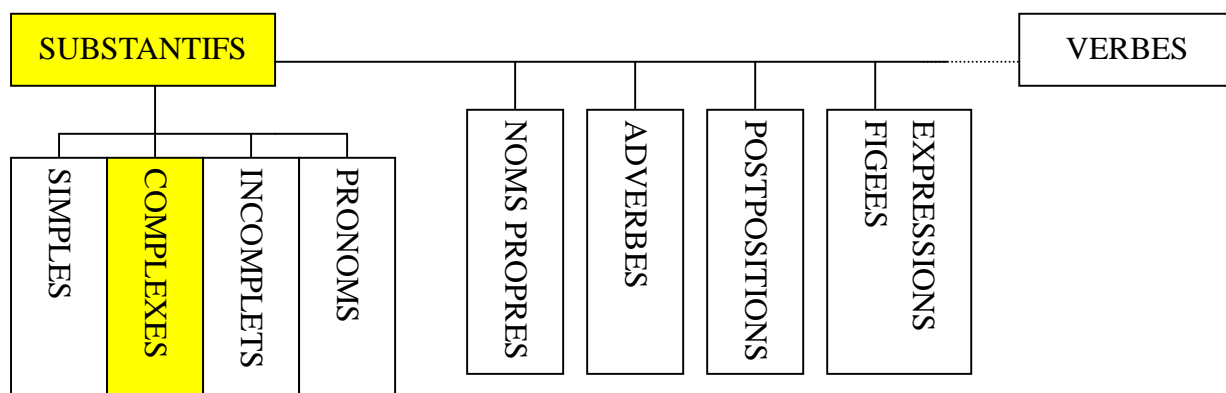


Figure 3. Structure du dictionnaire des substantifs du *Projet Sejong 21*

Selon les rapports du *Projet Sejong 21* (1998, 1999, 2000, 2001), on décrit dans un dictionnaire électronique toutes les informations sur les éléments de mots. Le dictionnaire des substantifs complexes (grisé) couvre les substantifs dérivés et composés (cf. section 1-1, pp. 8-9). Pour les éléments sino-coréens, les mots simples sont définis comme étant les unités autonomes minimales. En général, une séquence de deux idéogrammes est une unité autonome. Par exemple, *wi-in* (grand-homme = grand homme), qui est constitué de deux idéogrammes, est un nom simple, et non un nom composé. Regardons quelques exemples de substantifs complexes tirés du rapport du *Projet Sejong 21* (2001) :

- (15) a. *chamga-ja* (*N-Sf*) (participation - personne = participant)
- b. *jiha-cheol-yeok* (*(N-Sf)-N*) ((sous-sol - fer = métro) - station = station de métro)
- c. *cheonyeon-ginyeom-mul* (*N-N-Sf*) (nature - commémoration - objet = sites naturels protégés et espèces protégées)
- d. *hwang-so-gojip* (*(N-N)-N*) ((jaune - bœuf = taureau) - entêtement = entêtement forcené)
- e. *hoebok-gi-bogyun-ja* (*(N-Sf)-(N-Sf)*) ((guérison - période = convalescence ; porteur de microbes - personne = porteur de microbes) = porteur de microbes en convalescence)

(15a) est un nom dérivé et (15b) - (15e) sont des noms composés. (15c) est un nom composé du type « base + nom dérivé » (*N-(N-Sf)*) ou « nom composé + affixe » (*(N-N)-Sf*). (15d) est doublement composé : *hwang-so* (jaune - bœuf = taureau) est déjà un

nom composé. Selon le rapport du *Projet Sejong 21*, toutes les informations sur la composition sont bien décrites²⁶. De plus, les informations linguistiques comme les variantes orthographiques, l'origine, les caractères chinois pour les éléments sino-coréens, les caractères latins pour les éléments empruntés, les synonymes, les hyperonymes, les hyponymes, la traduction vers l'anglais, etc. sont aussi codés. Voici l'exemple *hwang-so-gojip* « entêtement forcené » dans un dictionnaire électronique :

● 황소고집

```
<entry>
  %writer=[]
  %dic=[]
  %crd=[]
  %mdd=[]
  %note=[]
  <toplevel>
    @form=[황소고집]
    @pos=[nn]
    @see=[]
    <morph_a>
      @var=[xs={ }; xd={ }; xx={ }]
      @abb=[]
      @lng=[]
      @str=[{n(황)+n(소)}+n(고집)]
      @org=[ko+si(황소+固執)]
    </morph_a>
    <morph_b>
      @hom=[]
      @der=[]
      @comp=[]
      @metc=[]
    </morph_b>
    <froz>
      @idnp=[]
      @idna=[]
      @idnv=[]
      @idda=[]
      @prv=[]
      @idetc=[]
    </froz>
  </toplevel>
  <sense n=1>
    @eg=[철수가 황소고집을 부리다]
    @trans=[stubbornness]
    @domain=[]
    @reg=[]
    @con=[]
    @curs=[U]
```

²⁶ Pour le moment, 40000 substantifs complexes sont recensés dont 500 décrits en détail, et 20000 entrées seront ajoutées.

```

@sem=[ 감정 ]
<lr>
    @syn=[ 소고집; 쇠고집 ]
    @ant=[]
    @hyper=[ 성격 ]
    @hypo=[]
    @coord=[]
    @holo=[]
    @mero=[]
    @rel=[]
</lr>
<synt_a>
    @cl=[ uni=; grp=; div=; qnt= ]
    @prt=[]
    @av=[]
    @ds=[]
</synt_a>
<synt_b>
    @comb_aj=[]
    @magn=[]
    @comb_v=[ ~을 부리다 ]
    @comb_n=[]
    @supv=[ ; cor_v={ } ]
    @max_n=[]
</synt_b>
<synt_c>
    @pre_d=[]
    @pre_n=[]
    @pre_s=[]
</synt_c>
</sense>
</entry>
</superEntry>
<superEntry>

```

Quant au travail de Nam J.S (1994, 1997), son organisation des dictionnaires électroniques du coréen se réfère au système *DELA* du LADL. Ce système est réparti en cinq dictionnaires : des mots simples, des mots affixés, des mots composés, des noms propres et des noms techniques (cf. figure 3, p. 30).

Le point de vue de Nam J.S sur la formation des noms diffère un peu de celui des linguistes traditionnels que nous avons exposé dans la section 1-1. Elle simplifie la présentation de la formation des noms en utilisant la notion d'autonomie. Un nom simple est un élément autonome et un nom dérivé est une combinaison entre des éléments autonomes et des éléments non autonomes. Un préfixe est défini comme un élément non autonome s'ajoutant en tête d'un nom et un suffixe comme un élément non autonome s'ajoutant à la fin d'un mot. Un mot composé est une composition d'éléments autonomes. La distinction entre noms dérivés et composés est donc effectuée selon le

nombre d'éléments autonomes : dans un nom composé le nombre d'éléments autonomes est égal ou supérieur à 2. Quant aux combinaisons d'éléments non autonomes comme *Pf-Rx*, *Rx-Sf* ou *Pf-Sf*, elle les considère comme des noms simples²⁷.

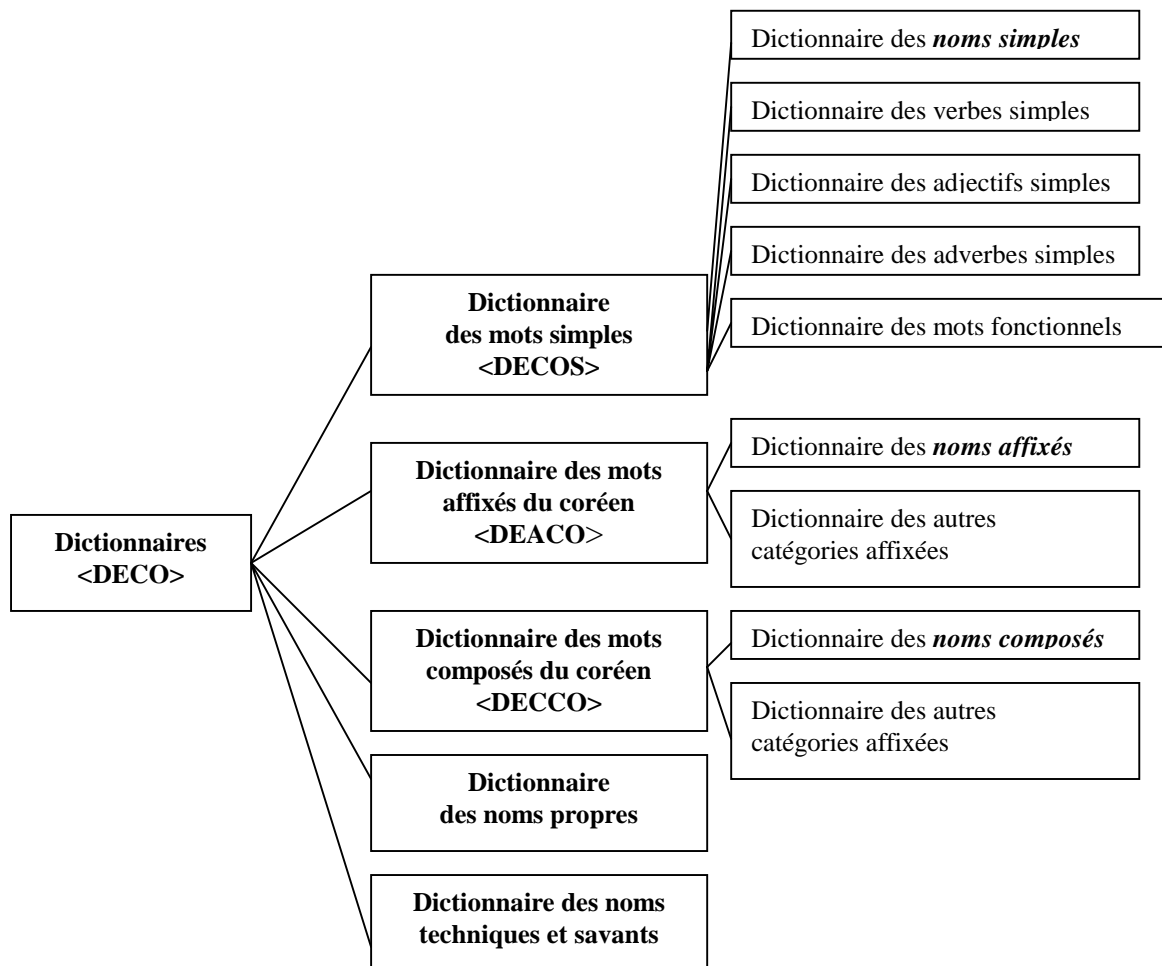


Figure 3. Ensemble des dictionnaires *DECO*²⁸

Elle a recensé 15000 entrées de noms simples sans dérivation ni composition, 900 préfixes et 900 suffixes lexicaux. Elle introduit une typologie des structures des noms dérivés²⁹ (cf. figure 4) et des noms composés³⁰ (cf. figure 5).

²⁷ Elle ne s'intéresse pas à l'analyse en morphèmes des mots qui ne comportent aucune partie autonome (cf. Nam J.S (1994), p. 38).

²⁸ Nam J.S (1994), p. 76.

²⁹ Nam J.S (1994), p. 38.

³⁰ Nam J.S (1994), p. 59.



Figure 4. Graphe des structures des noms simples et des noms dérivés³¹

Citons ses exemples de nom simple (16a) et de noms dérivés (16b) - (16f):

- (16) a. *gaeul* (N) « automne »
 b. *jae-gyeolhap* (Pf-N) « re » « unification » = « réunification »
 c. *podo-ju* (N-Sf) « alcool » « raison » = « vin »
 d. *dong-seong-ae* (Pf-N-Sf) « même » « sexe » « amour » = « amour homosexuel »
 e. *ga-chul-ok* (Pf-Pf-N) « provisoire » « sortir » « prison » = « sortie provisoire de prison »
 f. *yeongli-bu-jang* (N-Sf-Sf) « comptabilité » « département » « responsable » = « responsable du département de comptabilité »

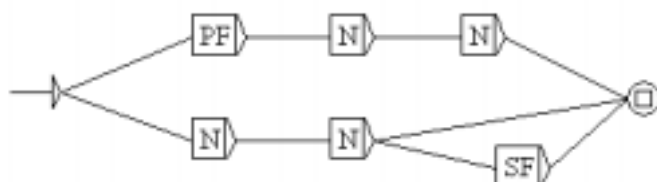


Figure 5. Graphe des structures des noms composés³²

Les exemples suivants sont des noms composés :

- (17) a. *gusul-ttam* (N-N) « perle » « sueur » = « goutte de sueur »
 b. *ban-dokjae-tujaeng* (Pf-N-N) « anti » « dictature » « révolte » = « révolte anti-dictature »
 c. *ganho-bojo-won* (N-N-Sf) « infirmier » « assistance » « personne » = « assistant

³¹ Nam J.S (1994), p.40.

³² Nam J.S (1994), p.40.

d'infirmier »

d. *gaundet-son-galak* (N-N-N) « centre » « main » « doigt » = « majeur »

Elle observe des séquences nominales comportant des éléments non autonomes sino-coréens à deux syllabes et introduit le terme de *pseudo-noms* (PN)³³ pour cette notion d'éléments non autonomes sino-coréens à deux syllabes. Elle recense 262 pseudo-noms :

- (18) a. *minju jeongchi* (PN N) « démocratique » « société » = « société démocratique »
 b. *jamjeong jochi* (PN N) « provisoire » « décision » = « décision provisoire »
 c. *jeongi jiphoe* (PN N) « régulier » « réunion » = « réunion régulière »
 d. *janggi chilyo* (PN N) « longue durée » « traitement » = « traitement de longue durée »
 e. *gongsan jeonggwon* (PN N) « communiste » « régime » = « régime communiste »

Elle propose la structure suivante pour la représentation des noms dans un dictionnaire électronique (Nam J.S (1997), p.13) :

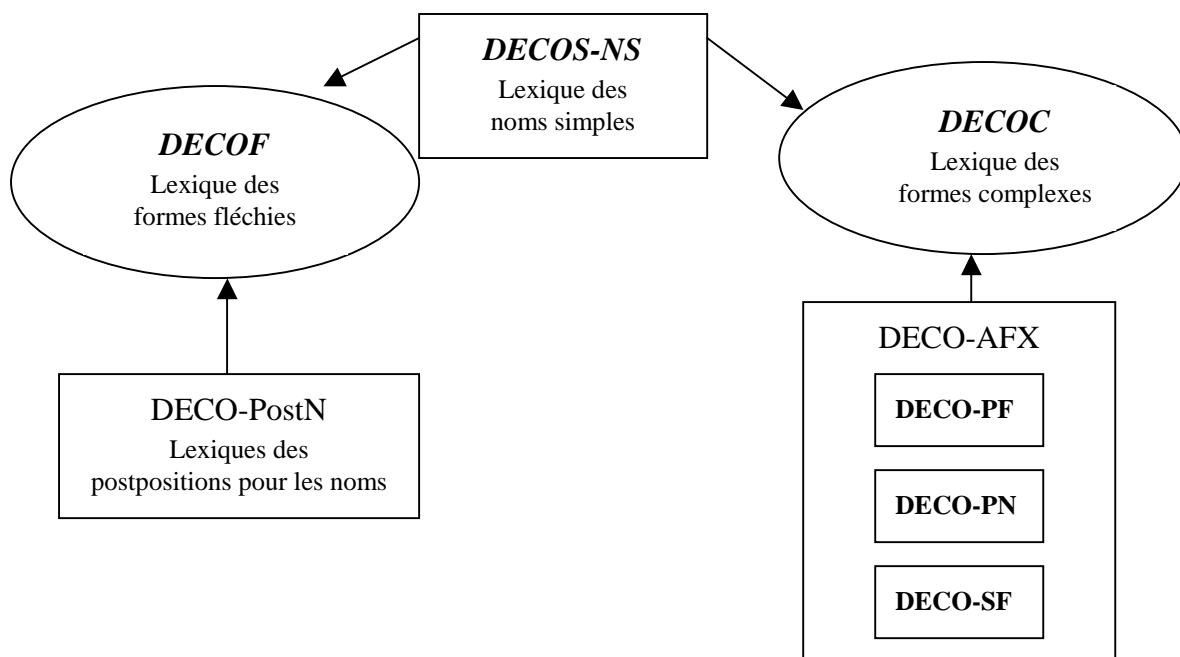


Figure 4. Système lexical des noms en coréen

³³ Le statut des *pseudo-noms* est intermédiaire entre celui des préfixes et celui des noms simples. Nous reviendrons à la notion *pseudo-noms* dans le chapitre 2 (2-1. La définition de *pseudo-mot X*).

Même si elle définit les mots dérivés et composés autrement que la tradition linguistique, il nous semble que le point de vue de Nam J.S (1997) se fonde sur la formation des mots pour la structuration du dictionnaire électronique.

Nous avons examiné deux dictionnaires électroniques représentatifs des substantifs en coréen fondés sur la formation des mots. Même s'il est intéressant de codifier toutes les informations étymologiques sur la composition des mots, comme nous l'avons vu, il est très difficile de décider le statut morphologique de chaque élément sino-coréen monosyllabique. En fait, le vocabulaire sino-coréen constitue une grande partie du lexique du coréen et il est pratiquement impossible de prévoir toutes les compositions contenant des éléments sino-coréens.

Pour nous, un dictionnaire sans informations d'espacement entre éléments n'est pas exhaustif parce que nous constatons beaucoup de variations des noms composés liées à l'espacement. Le fait que les unités lexicales composés (ou figées) aient des variantes typographiques ou comportent des espaces entre leurs éléments est important pour reconnaître les unités lexicales dans les textes par la consultation algorithmique du dictionnaire électronique. Si on structure le dictionnaire électronique en noms simples, noms dérivés et noms composés au sens traditionnel, on n'en tire aucun intérêt algorithmique au point de vue de l'efficacité du traitement informatique. De plus, nous ne pouvons pas utiliser les informations sur la composition des mots pour la segmentation des mots inconnus, parce que les informations sur les éléments des mots (surtout sur les éléments sino-coréens) nous donnent trop d'ambiguïtés. Par exemple, une syllabe sino-coréenne *ga* écrite dans l'alphabet coréen peut correspondre à plusieurs idéogrammes chinois (賈, 稼, 嘉, 架, 佳, 暇, 伽, 街, 歌, 假, 加, 可, 價, 家, 傢, 咖, 耬, 哿, 嫁, 譌, 葭, 軻, 笳, 灾, 罌, 榎, 痂, 袈, 柯, 珂, 迦, 苛, 茄, et 駕). C'est pourquoi les ingénieurs dans le domaine de l'informatique linguistique évitent que la segmentation des noms inconnus ne forme un segment monosyllabique³⁴ et considèrent les éléments monosyllabiques comme des affixes. Cette classification linguistique n'a d'intérêt que pour l'étymologie des éléments composés ou peut-être l'organisation de la maintenance du dictionnaire.

4. Conclusion

Nous avons examiné plusieurs points de vue principaux sur les noms composés

³⁴ cf. Voir les quatre règles de décomposition dans la section 2-2-1, p. 22.

du coréen. En fonction des objectifs pour lesquels le dictionnaire est conçu, la structure ou les informations à codifier peuvent être différentes. Notre but n'est pas d'analyser la composition des noms, mais de construire un dictionnaire électronique des noms composés pour plusieurs domaines appliqués du traitement automatique des langues naturelles. Comment peut-on délimiter les noms composés du coréen ? La délimitation des noms composés par opposition aux groupes nominaux libres est un problème fondamental. Devant la construction du dictionnaire électronique, nous nous confrontons à un problème très délicat : d'une part, il faut distinguer les notions syntaxiques, distributionnelles et sémantiques et d'autre part considérer l'usage typographique pour l'identification automatique. Les difficultés viennent du fait que les relations entre les notions linguistiques et typographiques sont complexes. Les combinaisons libres dans lesquelles entrent les mots coréens (surtout les combinaisons de mots sino-coréens) et l'espacement irrégulier posent de grandes difficultés dans la délimitation non seulement des séquences nominales figées vis-à-vis des séquences nominales libres mais aussi entre les différentes catégories de séquences nominales figées. Lors de la construction du dictionnaire électronique du coréen, nous devons faire des choix spécifiques à l'intérieur de la méthodologie du LADL. Nous allons introduire dans le chapitre 2 une nouvelle délimitation des noms composés pour la construction du dictionnaire électronique des noms composés.

Chapitre 2. Les séquences nominales figées dans un dictionnaire électronique

Dans ce chapitre nous allons voir comment nous délimitons les noms composés dans un dictionnaire électronique. Nous commençons à introduire la notion de figement dans la section 1. La section 2 est consacrée aux critères linguistiques de figement des séquences nominales : l'opacité sémantique, les restrictions distributionnelles, les paramètres syntaxiques comme le blocage des propriétés transformationnelles, les restrictions sur l'insertion, l'acceptabilité de N_1N_2 est (*un* N_2 + *un type de* N_2) et les séquences nominales figées de structure XN . Après avoir examiné les variations typographiques et les conventions typographiques, nous en déduisons une classification des séquences nominales figées dans la section 3.

1. La notion de figement : les séquences nominales figées et libres

La propriété principale du figement est que les locuteurs ne peuvent pas créer librement les séquences figées avec les règles de grammaire coréenne¹. Ces formes sont institutionnalisées et appartiennent au lexique coréen. Pour construire le dictionnaire électronique des séquences nominales figées, tout d'abord, il est indispensable de distinguer les séquences figées des séquences libres.

Les séquences libres sont celles qui obéissent simplement aux règles combinatoires générales de la langue, et les séquences figées, par définition, celles qui ne sont pas tout à fait libres dans la combinaison syntaxique et sémantique de leurs éléments. Le sens des séquences libres peut être déduit de celui de leurs éléments. Or, il n'est pas facile de distinguer selon des critères précis les séquences nominales figées et les séquences nominales libres.

En coréen, les séquences composées de type NN^+ , c'est-à-dire NN , NNN , $NNNN...$, constituent la grande majorité des composés nominaux. Comme nous l'avons vu dans le chapitre 1, les combinaisons de noms sont beaucoup plus libres en coréen qu'en français et sont en nombre théoriquement infini. Reprenons l'exemple (12b) dans la section 2-1 du chapitre 1 et considérons les combinaisons de noms :

- (1) a. *gukhoe* (N) « l'Assemblée nationale »
b. *gukhoe uijang* (N N) « l'Assemblée nationale » « président » =

¹ cf. G.Gross (1996), M.Silberstein (1993) et Lee H.J (1995).

- « président de l'Assemblée nationale »
- c. *gukhoe uijang seongeo* (N N N)
 « l'Assemblée nationale » « président » « élection » =
 « élection du président de l'Assemblée nationale »
- d. *gukhoe uijang seongeo gaesi* (N N N N)
 « l'Assemblée nationale » « président » « élection » « commencement » =
 « commencement de l'élection du président de l'Assemblée nationale »

L'exemple (1b) est une séquence nominale figée car elle est institutionnalisée et le sens n'est pas tout à fait transparent. En revanche, les séquences (1c) et (1d) sont des séquences nominales libres contenant la séquence nominale figée *gukhoe uijang*. (1c) obéit aux règles de la formation de groupes nominaux du coréen à partir d'une phrase $N_0 W N_1 (= : \textit{gukhoe uijang})\text{-Acc } N_{pred} (= : \textit{seongeo})\text{-hada}^2$. (1d) y obéit à partir d'une phrase $N_0 W N_1 (= : \textit{gukhoe uijang seongeo})\text{-Acc } N_{pred_2} (= : \textit{gaesi})\text{-Acc hada}$. Elles sont constituées à partir de la combinaison libre de *gukhoe uijang* avec *seongeo* ou de *gukhoe uijang seongeo* avec *gaesi*³. En fonction de leur degré de figement, ces séquences forment un continuum entre les séquences nominales figées et les séquences nominales libres. Toutefois, nous essayons d'établir les critères de figement dans la section 2.

Puisque notre but n'est pas de traiter quelques centaines ou milliers de phrases mais des textes, nous devons inclure les séquences nominales figées avec la plus grande couverture. Nous utilisons donc le terme *séquences nominales figées* au sens large, qui couvrent les noms composés proprement coréens (2a), les collocations nominales (2b), les expressions idiomatiques nominales (2c) au sens traditionnel et les séquences nominales figées sino-coréennes (2d). Pour éviter des ambiguïtés avec la terminologie traditionnelle, nous parlons de *séquences nominales figées* au lieu de noms composés. Les conventions typographiques sont variables selon les types de séquences nominales figées (pour les conventions typographiques, voir la section 3 de ce chapitre). Voici plusieurs exemples de séquences nominales figées.

- (2) a. *kkot-galu* (N-N) « fleur » « poudre » = « pollen »
 b. *bingsan-ui ilgak* (N-Gén N) « iceberg-Gén » « coin » = « la partie émergée »

² Cette construction à verbe support –*hada* est équivalente à $N_0 W N_1 N_{pred}\text{-Acc hada}$ et à $N_0 W N_1\text{-Acc } N_{pred}\text{-Acc hada}$.

³ *seongeo* « élection » et *gaesi* « commencement » sont des noms prédicatifs. Nous y reviendrons dans le chapitre 3-7 dans lequel nous définirons les noms prédicatifs.

de l'iceberg »

saepalga-n geojinmal (A-Sfd N) « être rouge foncé-Sfd » « mensonge »

= « gros mensonge »

c. *ko mut-eun don* (N V-Sfd N) « morve » « s'attacher à-Sfd » « argent »

= « argent de poche des enfants »

kkul meok-eun beongeoli (N V-Sfd N) « miel » « manger-Sfd » « sourd-muet »

= « être muet qui a mangé le miel ; personne qui ne peut pas exprimer ce qu'elle ressent »

sink-eun juk meok-gi (V-Sfd N N) « se refroidir-Sfd » « bouillie de riz »

« manger-Sfn » = « travail très facile à faire »

d. *gajok gyehoek* (N N) « famille » « projet » = « contrôle des naissances »

chuli soseol (N N) « déduction » « roman » = « roman policier »

tongsok soseol (Rx N) « populaire » « roman » = « roman populaire »

Le figement des noms composés au sens traditionnel (2a), des collocations et des expressions idiomatiques ((2b) - (2c)) a fait l'objet principal d'études dans la tradition linguistique. Mais le figement des séquences sino-coréennes est peu étudié. Notre travail sur les critères de figement des séquences nominales figées n'est donc pas limité aux noms composés ni aux expressions idiomatiques au sens traditionnel et il est même plutôt focalisé sur les séquences nominales sino-coréennes comme (2d).

2. Les critères de figement

2-1. Opacité sémantique

Ce critère lié au sens est un critère traditionnel de définition non seulement des noms composés mais aussi des expressions idiomatiques. C'est un critère fondamental pour les séquences nominales figées. Si le sens des séquences nominales ne peut pas être calculé à partir du sens de chaque élément, et si le résultat voulu du traitement des séquences nominales n'est pas prévisible (par exemple, leur traduction ou leur inclusion dans un index), ils sont lexicalisés et il est nécessaire de les décrire dans un dictionnaire électronique.

Par exemple, *iut sachon* (NN) « voisinage - cousin = voisin familial comme un cousin », *don setak* (NN) « argent - lavage = blanchiment d'argent », *pali moksum* (NN) « mouche - vie = vie de peu de valeur », *saepalga-n geojinmal* (A-Sfd N) « rouge foncé-Sfd - mensonge = gros mensonge », etc. doivent être inclus dans un

dictionnaire électronique parce que chaque séquence est sémantiquement une unité lexicale.

2-2. Restrictions distributionnelles

La substitution de synonymes est très limitée dans les séquences nominales figées telles que (3) ci-dessous. La structure de (3) est *A-Sfd N* :

- (3) a. *saeppaalga-n geojinmal* « être rouge foncé-*Sfd* - mensonge = gros mensonge »
 b. **ppalga-n geojinmal* « être rouge-*Sfd* - mensonge = ? »
 c. **saeppalga-n yueonbieo* « être rouge-*Sfd* - rumeur fausse = ? »
 d. **saeppalga-n heowisasil* « être rouge-*Sfd* - fait faux = ? » (Lee I.S, 1993)

La substitution de *ppalga-n* « être rouge » à *saeppalga-n* « être rouge foncé » dans (3b) et la substitution de synonymes tels que *yueonbieo* « rumeur fausse » ou *heowisasil* « fait faux » dans (3c) - (3d) sont exclues. Les substitutions de (4) sont similaires :

- (4) a. *yeolgwang-ui dogani* « enthousiasme-*Gén* - creuset = enthousiasme délirant »
 b. **yeolgwang-ui ttukbaegi* « enthousiasme-*Gén* - cuvette de terre = ? »
 c. **yeolgwang-ui daejeop* « enthousiasme-*Gén* - bol à potage = ? »
 d. **yeolgwang-ui sabal* « enthousiasme-*Gén* - cuvette de porcelaine = ? »
 (Kim J.H, 2000)

Dans (4), la substitution de synonymes tels que *ttukbaegi* « cuvette de terre », *daejeop* « bol à potage » ou *sabal* « cuvette de porcelaine » à *dogani* « creuset » est exclue.

2-3. Paramètres syntaxiques

2-3-1. Blocage des propriétés transformationnelles

Comme on le sait (M. Gross, 1986b), des restrictions à certaines propriétés transformationnelles existent même dans des séquences libres et aussi certaines transformations sont applicables à des séquences figées.

- (5) a. *chingu-(ui + E)(chaek + gabang + ot) (N-Gén N = N N)*
 (ami-(*Gén* + *E*) - (livre + sac + costume) = (livre + sac + costume) de l'ami)

- b. *yeolgwang*-(*ui* + **E*)(*bunwigi* + *chukje* + *pati*) (*N-Gén N* ; * *N N*)
 (enthousiasme fort -(*Gén* + **E*) (atmosphère + festival + fête) = (atmosphère + festival + fête) endiable(e))
- b'. *yeolgwang*-(*ui*+ **E*) *dogani* (enthousiasme-*Gén* - creuset = enthousiasme délirant)

Les exemples de (5a) - (5b) sont des séquences nominales libres contrairement à la séquence nominale figée (5b'). Cependant, les exemples libres (5b) et l'exemple figé (5b') ne permettent pas l'omission de la postposition de génitif. L'exemple suivant est le cas de la nominalisation d'une locution verbale (6a) en une locution nominale (6b) :

- (6) a. *Ina-neun* *geu* *sasil-lo* *gaseum-eul* *al-at-da*
 Ina-**Aux** ce fait-**Postp.lo** cœur-**Acc** (être malade)-**Mpas-St**
 (Ina a le cœur brisé à cause de ce fait)
- b. *Ina-neun* *geu* *sasil-lo* *gaseum ali-eul* *ha-yeot-da*
 Ina-**Aux** ce fait-**Postp.lo** cœur brisé-**Acc** faire-**Mpas-St**
 (Ina a le cœur brisé à cause de ce fait)

Par conséquent, la possibilité d'omission de la postposition de génitif et la nominalisation ne sont pas des critères suffisants de liberté.

Dans la structure *A-Sfd N*, la non-prédicativité de l'adjectif est une restriction plus opératoire pour distinguer les séquences nominales figées.

- (7) a. *Minu-ga* *saepalga-n* *moja-leul* *sse-ot-da*
 Minu-**Nmtf** rouge foncé-**Sfd** chapeau-**Acc** avoir-**Mpas-St**
 (Minu portait un chapeau rouge foncé)
- b. *moja-ga* *saepalga-t-da*
 chapeau-**Nmtf** (être rouge foncé)-**Mpas-St**
 (Le chapeau était rouge foncé)

Dans une séquence nominale libre telle que (7), un adjectif peut naturellement apparaître comme un modifieur d'un nom postposé ou comme un prédicat. Cependant, les séquences nominales figées ne permettent pas la prédicativité d'un adjectif :

- (8) a. *Minu-ga* *ina-ege* *saepalga-n* *geojinmal-eul* *ha-yeot-da*
 Minu-**Nmtf** Ina-**Postp.ege** rouge foncé-**Sfd** mensonge-**Acc** faire-**Mpas-St**

(Minu a fait un gros mensonge à Ina)

- b. * *geojinmal-i saeppalga-t-da*
 * mensonge-**Nmtf** (être rouge foncé)-**Mpas-St**
 (Le mensonge est rouge foncé)

Nous observons également qu'un complément libre d'un prédicat se comporte comme un constituant pour les transformations syntaxiques comme la pronominalisation et l'extraction. Par exemple :

- (9) a. *Minu-ga eoneohak gongbu-leul ha-yeot-da*
 Minu-**Nmtf** linguistique étude-**Acc** faire-**Mpas-St**
 (Minu a fait des études de linguistique)
- b. *Minu-ga mueot-eul gongbuha-yeot-ni?*
 Minu-**Nmtf** quoi-**Acc** étudier-**Mpas-Sti**
 (Qu'est-ce que Minu a étudié?)
- c. *Minu-ga gongbuha-n geos-eun eoneohak-i-da*
 Minu-**Nmtf** étudier-**Sfd** **Ncomp-Aux** linguistique-être-**St**
 (C'est la linguistique que Minu a étudiée)
- (10) a. *Minu-wa Ina-ga gajok gyehoek-eul ha-yeot-da*
 Minu-**Conj** Ina-**Nmtf** famille projet-**Acc** faire-**Mpas-St**
 (Minu et Ina ont pratiqué le contrôle des naissances)
- b. **Minu-wa Ina-ga mueot-eul gyehoekha-yeot-ni?*
 *Minu-**Conj** Ina-**Nmtf** quoi-**Acc** faire le projet-**Mpas-Sti**
 (Qu'est-ce que Minu et Ina ont projeté?)
- c. **Minu-wa Ina-ga gyehoekha-n geos-eun gajok-i-da*
 *Minu-**Conj** Ina-**Nmtf** fait le projet-**Sfd** **Ncomp-Aux** famille-être-**St**
 (C'est la famille que Minu et Ina ont projetée)

Les exemples (9) et (10) sont des structures N_1N_2 , où les N_2 sont des noms prédicatifs et les N_1 sont les objets compléments directs des N_2 . Dans les exemples libres (9b) - (9c), la pronominalisation et l'extraction interrogatives sont acceptables, tandis que ces transformations sont bloquées dans les exemples figés (10b) - (10c).

2-3-2. Restrictions sur les insertions

Les séquences nominales figées ne permettent aucune insertion d'adverbes

d'intensité tels que *aju* (très), *mani* (beaucoup), etc. comme dans (11b), où l'insertion d'un adverbe élimine le sens figé et sélectionne son sens littéral.

(11) a. *sik-eun juk meok-gi* (V-Sfd N N)

(se refroidir-Sfd - bouillie de riz - manger-Sfn = travail très facile à faire)

b. **sik-eun juk mani meokgi* (V-Sfd N Adv N)

(se refroidir-Sfd - bouillie de riz - beaucoup - manger-Sfn = action de manger beaucoup de bouillie de riz refroidie ; sens littéral)

Il n'est possible de qualifier les séquences nominales figées que dans leur totalité, comme dans (11c), où *wanjeonhi* (complètement) qualifie la séquence *sik-eun juk meokgi*, non pas une partie comme *sik-eun* :

c. *wanjeonhi* [*sik-eun juk meokgi*] (Adv V-Sfd N N)(complètement - se refroidir-Sfd - manger de la bouillie de riz = travail très facile à faire)

2-3-3. Acceptabilité de N_1N_2 est (un N_2 + un type de N_2)

G. Gross (1986) et J. Baptista (1995) présentent un critère intéressant pour l'étude des séquences nominales figées de structure NA en français et NN en portugais :

(12) a. **Um banho-maria é (um + um tipo de) banho* (N_1N_2)

(*Un bain-marie est (un + un type de) bain))

b. *Uma cobra-capelo é (uma + um tipo de) cobra* (N_1N_2)

(= Un serpent à sonnette est (un serpent + un type de serpent))

b'. *Le vin (rouge + rosé + blanc) est (un vin + un type de vin)* (NA)

Le sens des séquences nominales de (12a) ne peut pas être déduit du sens de chaque élément. Contrairement à (12a), les N_1 et N dans (12b) et (12b') gardent leur sens ; N_2 et A spécifient le sens de N_1 et de N . N_2 est un sous-classificateur de N_1 (12b) et A de N (12b'). Les exemples (12a), (12b) et (12b') sont figés, mais (12a) est plus figé que (12b) ou (12b').

Ce critère de l'acceptabilité de N_1N_2 est (un N_2 + un type de N_2) est très utile pour la structure NN du coréen.

(13) a. *dongne byeongwon* (N_1N_2) « village » « hôpital » = « hôpital de village »

- b. *dongne byeongwon-eun* (*byeongwon* + **byeongwoen-ui han yuhyeong*)-i-da
hôpital village-**Aux** (hôpital + **hôpital-Gen* un type)-être-**St**
(Un hôpital de village est (un hôpital + **un type d'hôpital*))
- (14) a. *jeongsin byeongwon* (N_1N_2) « esprit » « hôpital » = « hôpital psychiatrique »
b. *jeongsin byeongwon-eun* (*byeongwon* + *byeongwoen-ui han yuhyeong*)-i-da
hôpital psychiatrique-**Aux** (hôpital + hôpital-**Gen** un type)-être-**St**
(Un hôpital psychiatrique est (un hôpital + un type d'hôpital))
- (15) a. *jigak daejang* (N_1N_2) « retard » « chef » = « personne qui est toujours en retard »
b. **jigak daejang-eun* (*daejang* + *daejang-ui han yuhyeong*)-i-da
** personne qui est toujours en retard-Aux* (chef + chef-**Gen** un type)-être-**St**
** (une personne qui est toujours en retard est (un chef + un type de chef))*

(13) est une séquence nominale libre, et (14) et (15) sont des séquences nominales figées. Tandis que le sens des séquences nominales (15a) ne peut pas être calculé à partir du sens de chaque élément, N_2 de (14a) garde son sens : dans (14), N_1 joue un rôle de sous-classificateur de N_2 ⁴. (15a) est plus figé que (14a).

2-4. Les séquences nominales figées à structure *XN*

Dans le chapitre 1, nous avons examiné la formation des mots dans la tradition linguistique. Rappelons que la composition directe entre un radical et un nom comme *Rx-N* est appelée *composition déviante* pour le vocabulaire proprement coréen (cf. section 1-2 du chapitre 1, pp. 12-14) : i.e. *heoltteok-gogae* (manière de haleter - col = col très escarpé à monter)⁵. Quant au vocabulaire sino-coréen, les noms sino-coréens sont en général formés par la combinaison de deux idéogrammes (cf. pp. 14-15) et ces formes dissyllabiques sont des noms autonomes en coréen. Or, nous constatons que certaines formes sino-coréennes dissyllabiques sont non autonomes. Comparons le cas des noms à deux syllabes (16a') à celui des formes non autonomes à deux syllabes (16b') et (16b'') :

- | | |
|-----------------------|---|
| (16) a. <i>*wi-ga</i> | (<i>*Rx-Nmtf</i> = <i>*grand-Nmtf</i>) |
| <i>*in-i</i> | (<i>*Rx-Nmtf</i> = <i>*homme-Nmt</i>) |
| a'. <i>wi-in-i</i> | ((<i>Rx-Rx</i> = <i>N</i>)- <i>Nmtf</i> = <i>grand homme-Nmtf</i>) |
| b. <i>*min-i</i> | (<i>*Rx-Nmtf</i> = <i>*peuple-Nmtf</i>) |

⁴ En coréen, le déterminant antéposé qualifie le déterminé postposé.

⁵ Dans l'usage, on peut mettre un blanc entre *Rx* et *N*.

- **ju-ga* (**Rx-Nmtf* = **maître-Nmtf*)
 b'. **minju-ga* (*(*Rx-Rx*)-*Nmtf* = **démocratique-Nmtf*)
 b''. *minju-seong-i* ((*Rx-Sfn*)-*Nmtf* = *N-Nmtf*)
 (démocratique - caractère = caractère démocratique)
minju jeongchi-ga ((*Rx N*)-*Nmtf* = *N-Nmtf*)
 (démocratique - politique = politique démocratique)

La forme dissyllabique *minju* n'apparaît en position de nom sujet ou complément que par combinaison avec un autre élément : le suffixe *-seong* (16b') ou le nom *jeongchi* (16b'').

La situation est similaire pour le vocabulaire emprunté à d'autres langues que le chinois. Il existe des formes empruntées qui ne sont pas des mots en coréen (par exemple, *balokeu* « baroque », *beseuteu* « best », *seleo* « seller », etc.), même si ces formes sont des mots en anglais. Les autres emprunts comme *ladio* « radio », *keopi* « café », *keomputeo* « ordinateur », etc. se comportent syntaxiquement comme des noms dans le système du coréen.

- (17) a. **balokeu-ga* (**Rx-Nmtf* = **baroque-Nmtf*)
 a'. *balokeu eumak-i* (*Rx N-Nmtf* = *N-Nmtf*)
 (baroque - musique = musique baroque)
 b. (*ladio + keopi + keomputeo*)-*ga* = (*N-Nmtf*) ((radio + café + ordinateur)-*Nmtf*)

Dans la section 2-4-1, nous allons définir comme *pseudo-mots* les formes non autonomes *Rx* qui peuvent entrer dans des séquences *Rx N* en permettant l'espace entre eux : les *Rx* proprement coréens, les formes non autonomes sino-coréennes dissyllabiques et les formes non autonomes empruntées. Puis, nous allons examiner dans la section 2-4-2 leurs propriétés spécifiques, qui montrent bien leur figement syntaxique.

2-4-1. La définition des *pseudo-mots*

Avant de définir les *pseudo-mots*, qui sont un type de *Rx* que nous notons *X*, revenons à la notion de *pseudo-nom* (*PN*)⁶ de Nam J.S (1994)⁷ qui est plus ou moins

⁶ cf. Le terme *pseudo-substantif* est utilisé par Han S.H (2000, note 29, p. 24). Elle a remarqué que quelques formes non autonomes *gwajung* « lourd », *teuksu* « spécial », *jeonmu* « totalement absent », etc. ont été considérées comme des noms prédicatifs dans Nam J.S (1996). Pour Han S.H, ces formes ne sont

reliée à notre notion de pseudo-mot. Rappelons qu'elle a nommé *pseudo-noms* les éléments non autonomes sino-coréens à deux syllabes.

- (18) a. *gangyeol-mi* *PN-N* « net » « beauté » = « beauté de netteté »
 gangyeol-che *PN-SF* « net » « style » = « style de netteté »
 geonmang-jeung *PN-SF*⁸ « manque de mémoire » « symptôme » = « amnésie »
 gukmun-hak *PN-SF* « littérature coréenne » « science » = « littérature coréenne »
 heuksa-byeong *PN-N* « peste » « maladie » = « peste »
- b. *ganjeop seongeo* *PN N* « indirect » « suffrage » = « suffrage indirect »
 gangbap gwannyeom *PN N* « obsédant » « idée » = « obsession »
 gukje sahoe *PN N* « international » « société » = « société internationale »
 wonsi jonggyo *PN N* « primitif » « religion » = « religion primitive »

Nam J.S (1994) ne distingue pas les exemples de (18a) de ceux de (18b). Car, au point de vue de la formation des mots, il n'y a pas de différence entre (18a) et (18b), et les *PN* sont des éléments non autonomes sino-coréens dissyllabiques sans catégorie grammaticale, à savoir les *Rx* sino-coréens dissyllabiques. Mais nous constatons quelques différences entre les deux séries d'exemples. Dans (18a), le deuxième élément est monosyllabique, on ne peut pas déterminer précisément le statut morphologique des éléments sino-coréens⁹, et les structures internes des séquences sino-coréennes peuvent être analysées de plusieurs façons. Par exemple, la forme *guk-mun-hak* « littérature coréenne » (18a) qui est constituée de trois éléments sino-coréens peut être analysée selon les points de vue au moins de trois façons (19a) - (19c) :

pas des noms, mais des *pseudo-substantifs* parce que ces formes en formant des verbes statiques par suffixation en *-hada* n'apparaissent jamais en position de nom sujet ou complément. Pour nous, ces formes sont des *Rx* dissyllabiques et ne peuvent que se combiner avec d'autres éléments pour former des mots :

<i>teuksu-ha-da</i>	(<i>Rx-Sfa-St</i> = <i>A</i>)	« spécial - être = être spécial »
<i>teuksu-seong</i>	(<i>Rx-Sfn</i> = <i>N</i>)	« spécial - <i>Sfn</i> = spécialité »
<i>teuksu gyoyuk</i>	(<i>Rx N</i> = <i>N</i>)	« spécial - éducation = éducation pour les enfants handicapés physiques ou mentaux »

⁷ cf. voir la section 3 du chapitre 1.

⁸ Puisque pour elle, les éléments non autonomes sont des affixes, elle a considéré la forme non autonome *jeung* (syntôme) comme un suffixe. Mais pour nous, c'est un *Rx*.

⁹ Comme nous l'avons vu dans la section 1-3 du chapitre 1, nous nous heurtons à une difficulté de délimitation entre radical, préfixe et déterminant en position de premier élément du mot, et entre radical, suffixe et nom incomplet en position de dernier élément du mot, même avec l'aide de critères.

- (19) a. *guk-mun-hak* ($Rx-Rx-Rx = N$) « pays » « littérature » « science » =
« littérature coréenne »
- b. *gukmun-hak* (PN (ou Rx)- $Sf = N$) « littérature coréenne » « science » =
« littérature coréenne »
- c. *guk-munhak* ($Pf-N = N$) « pays » « littérature » = « littérature coréenne »

L'analyse interne (19b) de Nam J.S (1994) est une analyse possible. Quant aux conventions typographiques, lorsque les *Rx* dissyllabiques se combinent avec des éléments monosyllabiques (affixe, radical ou nom), les formes composées sont typographiquement soudées. Au contraire, lorsque les *Rx* dissyllabiques constituent des séquences nominales avec des noms à plus de deux syllabes, l'espacement entre les *Rx* et les noms est facultatif. On permet donc de laisser l'espace entre ces radicaux et les noms en (18b), tandis que l'on ne laisse pas de blanc en (18a)¹⁰.

Nous ne voyons guère d'intérêt ni informatique ni linguistique à analyser la structure interne des séquences nominales sino-coréennes obligatoirement soudées comme *Rx-N*, *Rx-Rx*, *Rx-Sf*, etc. : informatiquement ces séquences nécessitent le même algorithme d'identification et linguistiquement leurs structures internes peuvent être analysées de plusieurs façons. Nous proposons donc de ne pas faire l'analyse interne des séquences telles que (18a) et de les considérer comme des séquences nominales figées soudées (ou comme des mots dérivés ou composés). Par contre, l'analyse de la structure interne des exemples (18b) est importante. Linguistiquement, la structure interne des exemples (18b) est analysable de façon consensuelle (Nam J. S, 1994) et les séquences *PN N* de (18b) ont quelques propriétés spécifiques qui montrent bien leur figement syntaxique (cf. section 2-4-2). Puisque les *PN* (*Rx* dissyllabiques) de (18b) entrent dans les séquences nominales figées en permettant l'espace, informatiquement, leur identification automatique demande un algorithme différent par rapport à celle des séquences obligatoirement soudées comme (18a).

Par conséquent, nous nous intéressons aux *Rx* de (18b), et nous les appellerons *pseudo-mots*. Nous évitons les appellations *pseudo-noms* (*PN*) et *pseudo-substantifs* parce qu'elles lient ces éléments aux substantifs. Si les *Rx* se combinent avec des suffixes de nominalisation, ce sont des noms. Mais s'ils se combinent avec un suffixe de dérivation adjectivale, ce sont des adjectifs. Les catégories grammaticales des séquences constituées d'un *Rx* et d'autres éléments sont tout à fait dépendantes de ces derniers, mais non de *Rx*. Reprenons l'exemple de *gangyeol* dans (18a) :

¹⁰ Il nous semble que la possibilité d'insertion de l'espace entre les éléments fait souvent confondre ces radicaux avec des noms.

- (20) a. *gangyeol-mi* ($Rx-N = N$) « net » « beauté » = « beauté de netteté »
 b. *gangyeol-seong* ($Rx-Sfn = N$) « net » « caractère » = « netteté »
 c. *gangyeol-ha-da* ($Rx-Sfa-St = A$) « être net »

Nous appelons *pseudo-mots* et nous notons X , les radicaux Rx qui satisfont à la fois les trois conditions ci-dessous :

- (i) ils ont une ressemblance morphologique avec des mots, mais ce sont des éléments non autonomes, et ils ont un sens plein
- (ii) ils se combinent directement avec des noms sans aucun suffixe de dérivation syntaxique
- (iii) lorsqu'ils entrent dans la séquence $Rx N$, l'espacement est facultatif dans l'usage ou dans les dictionnaires éditoriaux

Par exemple, si un Rx ne se combine qu'avec des noms monosyllabiques, ce n'est pas un X :

- (21) a. *heuksa-byeong* ($Rx-N = N$) « peste » « maladie » = « peste »
 b. *muchae-saek* ($Rx-N = N$) « incolore » « couleur » = « certaines couleurs comme le blanc, le noir, le gris, etc. »

Au niveau de la formation des mots, les Rx dans (22) sont tout à fait similaires. Cependant, les Rx en (22a) et (22b) ne sont pas des pseudo-mots, alors que (22a') et (22b') sont des pseudo-mots.

- (22) a. *gyohang-gok* ($Rx-N = N$) « symphonie » « pièce » = « symphonie »
 a'. *gyohang ^ akdan* ($Rx-N = X N$) « symphonie » « orchestre » = « orchestre symphonique »
 b. *sangi-gun* ($Rx-N = N$) « blessure » « armée » = « armée de mutilés de guerre »
 b'. *sangi ^ gunin* ($Rx-N = X N$) « blessure » « militaire » = « mutilé de guerre »

Dans les dictionnaires éditoriaux, les Rx dissyllabiques sont en général étiquetés comme

radicaux¹¹ (i.e. *ganglyeok* « puissant », *gangyeol* « net », etc.) ou noms (i.e. *wonsi* « primitif », *gyohyang* « symphonie », *minju* « démocratique », etc.)¹². Si les formes composées comme (21) contenant des *Rx* qui ne sont pas des pseudo-mots sont des entrées dans les dictionnaires éditoriaux, elles sont étiquetées comme noms (*N*) avec un tiret qui marque la composition des deux éléments ou sans tiret (i.e. *muchae-saek* / *muchaesae* (*N*) « couleur incolore »). Si les formes composées contenant des *X* sont enregistrées dans les dictionnaires éditoriaux, elles y sont étiquetées comme noms (*N*), mais l'insertion d'un blanc ou d'un tiret est très dépendante des dictionnaires éditoriaux (i.e. *minju-jeonchi* / *minju jeonchi* / *minju ^ jeonchi* (*N*) « politique démocratique »).

Les *X* ne correspondent pas exactement aux éléments non autonomes sino-coréens à deux syllabes de Nam J. S (1994). La majorité des *X* sont des éléments sino-coréens ou d'origine étrangère. Si ce sont des mots dans leur système d'origine, ils ont perdu leur autonomie dans le système du coréen, au contraire de noms comme *chaek* « livre », *mun* « porte », *ladio* « radio », *keopi* « coffee » ou *keomputeo* « computer », etc. Examinons divers exemples de *X* qui entrent dans la structure *X N*.

Les *X* sino-coréens à deux syllabes sont nombreux :

- (23) a. *tongsok ^ soseol* (*XN*) (populaire - roman = roman populaire)
 b. *yeongse ^ gieop* (*XN*) (petite envergure - entreprise = entreprise de petite envergure)
 c. *bulon ^ munseo* (*XN*) (séditieux - écrit = écrit séditieux)
 d. *minju ^ jeongchi* (*XN*) (démocratique - politique = politique démocratique)

Les exemples de (24) sont des cas un peu spéciaux où un nom autonome est combiné avec le préfixe *ban-* (anti-), et cette séquence (*Pf-N*) est un *X* non autonome à trois syllabes. A part cette structure interne, on ne voit aucune différence entre (23) et (24).

- (24) a. *ban-sahoe* (*PF-N = Rx = antisocial*)
 ban-sahoe-seong (*Rx-seong = N = caractère antisocial*)
 ban-sahoe-jeok (*Rx-jeok = N = avoir une tendance antisociale*)
 ban-sahoe ^ jipdan (*X N*)(*antisocial - communauté = communauté antisociale*)
 b. **bansahoe-leul* (**X-Acc = antisocial-Acc*)
 basahoe-seong-i (*((Rx-Sfn = N)-Nmtf = caractère antisocial-Nmtf)*)

¹¹ Si les *Rx* expriment un état statique, ils sont considérés comme des radicaux adjectivaux.

¹² L'étiquetage des *X* peut être différent selon les dictionnaires.

bansahoe ^ *jipdan-i* (X N-Nmtf) (communauté antisociale-Nmtf)

Les exemples de (25) sont des cas proprement coréens souvent ressentis comme des onomatopées.

- (25) a. *kkobulang* (Rx = courbé)
kkobulang-hada (Rx-hada ; A = être courbé)
kkobulang-geolida (Rx-geolida ; V = faire des méandres)
kkobulang-kkobulang (Rx-Rx ; Adv = d'apparence crépue ; adverbe expressif)
kkobulang ^ *neulgeuni* (XN) (courbé - vieux = vieillard qui a le dos courbé ;
vieillard très vieux)
kkobulang ^ *neulgeuni-ga*(XN-Nmtf) (vieillard très vieux-Nmtf)
- b. *sandeul* (Rx = souffle)
sandeul-geolida (Rx-geolida ; V = souffler doucement)
sandeul-sandeul (Rx-Rx ; Adv = comme un léger souffle de vent ;
adverbe expressif)
sandeul-sandeul-hada (Rx-Rx-hada = Adv-hada ; A = être comme un léger
souffle de vent)
sandeul ^ *balam* (X N ; N) (souffle - vent = vent doux ; brise)
sandeul ^ *balam-I* (X N-Nmtf) (brise-Nmtf)

Les éléments non autonomes Rx proprement coréens forment aussi des mots en se combinant avec d'autres éléments. Lorsque ces Rx se combinent avec des noms à plus d'une syllabe, l'espacement est facultatif (*kkobulang* ^ *neuleuni* ou *sandeul* ^ *balam*)¹³.

Il existe des X empruntés à d'autres langues que le chinois :

- (26) a. *balokeu* (Rx = baroque)
balokeu ^ *eumak* (X N) (baroque - musique = musique baroque)
balokeu ^ *eumak-i* (X N-Nmtf) (musique baroque-Nmtf)
- b. *beseuteu* (Rx = best ; en anglais)
seleo (Rx = seller ; en anglais)
beseuteu ^ *seleo* (X X) (best - seller = succès de librairie)

¹³ En fait, *kkobulang* ^ *neuleuni* ou *sandeul* ^ *balam* sont des noms composés au sens traditionnel, mais pour nous, *kkobulang* et *sandeul* sont des X parce que ces éléments satisfont les trois conditions voulues. Ces X ne sont pas nombreux.

beseuteu seleo-ga (X X-Nmtf)(best-seller-Nmtf)

En coréen, l'élément *seller* n'est pas autonome malgré son statut de nom en anglais. Comme en (26b), si les emprunts *beseuteu seleo* (X X) sont des noms composés en anglais lui-même, nous empruntons directement les formes entières X X.

2-4-2. Plusieurs types de XN

Nous allons examiner les propriétés syntaxiques qui montrent bien le figement syntaxique en coréen. Il y a trois types de Rx qui peuvent être un élément X dans XN à espacement facultatif ($Rx\ N = X\ \wedge\ N$) :

- Type A. $Rx (-jeokin + *-han + E) N$
- Type B. $Rx (*-jeokin + -han + E) N$
- Type C. $Rx (*-jeokin + *-han + E) N$

-jeok-ida et *-hada* dans *Rx-jeok-ida*¹⁴ (être comme *Rx*) et *Rx-hada* (être *Rx*) sont des suffixes de dérivation syntaxique et *-n* dans *-jeokin* et *-han* est un suffixe déterminatif.

Avant de discuter le Type A de XN , tout d’abord, regardons en détail le problème de la compatibilité¹⁵ entre Rx et $-jeok(in)$, $Rx-jeok$ et N , et $Rx-jeokin$ et N . Il nous semble que $-jeok$, qui est sino-coréen, est un suffixe de dérivation très productif. Cependant, il y a quelques incompatibilités : $-jeok$ est compatible avec les éléments sino-coréens et empruntés (noms ou Rx), et avec les noms propres (27a), mais il n’est pas compatible avec les éléments proprement coréens (27b) sauf les noms propres proprement coréens (27a’).

(27) a. *minju-jeok* *jeongchi* (*sc-jeok* *sc*) (*Rx-jeok* *N*)
 (démocratique - politique = politique démocratique)
saediseum-jeok *soseol* (*emp-jeok* *sc*) (*N-jeok* *N*)¹⁶
 (sadique - roman = roman sadique)

¹⁴ Nous avons considéré *Rx-jeok-ida* comme un adjectif. Au niveau de la formation, *Rx-jeok-ida* ainsi que *Rx-hada* sont des adjectifs dérivés. Puisque *Rx-jeok* « comme *Rx* » exprime un état statique et partage des propriétés syntaxico-sémantiques avec l’adjectif, ce *Rx* combinable avec *-jeok* est considéré comme un radical dit adjectif d’après Kim C.S (1996) et No M.H (1998). Nam J.S (1996, pp. 35-38) définit ces *Rx-jeok-ida* et *Rx-hada* comme des adjectifs simples au niveau syntaxico-sémantique.

¹⁵ cf. Seo J.S (1996), pp. 583-585.

¹⁶ *emp* indique un élément emprunté, *sc* un élément sino-coréen et *pc* un élément proprement coréen.

- maleukeuseu-jeok sago (emp-jeok sc) (N-jeok N)*
 (comme Marx - pensée = pensée marxienne)
- a'. *Kim Gu-jeok sago (pc-jeok sc) (N-jeok N)*
 (comme Kim-Gu - pensée = pensée apparentée à celle
 de Kim-Gu ; personnage historique coréen)
- b. **nala-jeok gyumo (*pc-jeok sc) (*N-jeok N)*
 (*national - envergure = envergure nationale)
- gukga-jeok gyumo (sc-jeok sc) (N-jeok N)*
 (national - envergure = envergure nationale)

Toutefois, tous les noms sino-coréens ne sont pas compatibles avec *-jeok* :

- (28) **namja-jeok (*sc-jeok)(*N-jeok = homme - avoir une tendance)*
**noin-jeok (*sc-jeok)(*N-jeok = vieux - avoir une tendance)*

En général, *-jeok* et *-jeokin* commutent¹⁷. Si le *N* est un nom sino-coréen comme en (29), l'ajout et l'effacement du morphème coréen *-in* sont acceptables.

- (29) a. *minju-jeok ingan (sc-jeok sc) (Rx-jeok N)*
 (démocratique- homme = homme démocratique)
- b. *minju-jeoki-n ingan (sc-jeok-i-n sc) ((Rx-jeoki)-n N = A-Sfd N)*
 (être démocratique-Sfd-homme = homme démocratique)

Dans *Rx-jeok N*, si le *N* déterminé par *Rx-jeok* est un nom proprement coréen, la forme sans le morphème proprement coréen *-in* dans cette séquence est peu naturelle, mais l'ajout de ce morphème est naturel.

- (30) a. **minju-jeok salam (*sc-jeok pc) (*Rx-jeok N)*
 (démocratique - homme = homme démocratique)
- b. *minju-jeokin salam (sc-jeokin pc) (A-Sfd N)*
 (être démocratique-Sfd - homme = homme démocratique)

Nous pouvons résumer les restrictions comme suit :

¹⁷ *Rx-jeok* et *Rx-jeokin* sont toujours commutables sauf un seul cas exceptionnel : les *Rx-jeok* sino-coréens ne déterminent pas des noms proprement coréens.

<i>Rx</i>	<i>N</i>	<i>Rx-jeokin</i> <i>N</i>	<i>Rx-jeok</i> <i>N</i>	<i>Rx N (= X N)</i>
sc/emp	sc/emp	sc/emp-jeokin sc/emp	sc/emp-jeok sc/emp	sc/emp sc/emp
	pc	sc/emp-jeokin pc	*sc/emp-jeok pc	sc/emp pc
pc	-	*pc-jeokin <i>N</i>	*pc-jeok <i>N</i>	*pc <i>N</i>

Figure 1. Compatibilité entre *Rx* (-jeok(in) + *-han+E) et le nom déterminé

Revenons aux exemples de Type A (= : *Rx* (-jeokin + *-han + *E*) *N*). Les exemples de (31a) sont des séquences nominales libres où *Rx-jeok* est prédicatif ((31a')). Mais l'omission du suffixe -jeokin de (31b) est interdite.

- (31) a. *tongsokjeok-i-n* (*chaek* + *japji* + *naeyong* + *sasil*) (*A-Sfd N*)
 (populaire-être-*Sfd* (livre + magazine + contenu + fait) = (livre + magazine + contenu + fait) populaire)
 a' (*chaek* + *japji* + *naeyong* + *sasil*)-*i* *tongsokjeok-i-da*
 (livre + magazine + contenu + fait) -**Nmtf** populaire-être-**St**
 (Le (livre + magazine + contenu + fait) est populaire)
 b. **tongsok* (*chaek* + *japji* + *naeyong* + *sasil*) (**X N*)
 *populaire (livre + magazine + contenu + fait)

Examinons les exemples de (32).

- (32) a. *tongsokjeok-i-n* *soseol* (*A-Sfd N*)
 (populaire-être-*Sfd* - roman = roman qui a de la popularité)
 a' *soseol-i* *tongsokjeok-i-da*
 roman-**Nmtf** populaire-être-**St**
 (Le roman est populaire)
 b. *tongsok* ^ *soseol* (*X N*)(populaire - roman = roman populaire)
 b' **soseol-i* *tongsok-i-da*
 *roman-**Nmtf** populaire-être-**St**

Au niveau syntaxique, *X* n'est pas prédicatif (32b'). Tandis que (32a) est une séquence nominale libre, (32b) est une séquences nominale figée. Même dans des séquences nominales figées de structure *XN* comme (32b), il nous semble que le sens de *XN* peut être déduit de celui de ses éléments. En fait, nous observons une différence sémantique

entre (32a) et (32b). Alors que le groupe nominal libre *tongsokjeoki-n soseol* (A-Sfd N) (32a) peut désigner spécifiquement n'importe quel roman qui a de la popularité, le nom figé *tongsok soseol* (XN) (32b) dénote génériquement un type ou genre de roman. Par conséquent, l'exemple (32b) figé entre dans la phrase *X N est un (N + type de N)*¹⁸ :

- (32) c. *tongsok soseol-eun* (soseol + soseol-ui han yuhyeong)-i-da
 roman populaire-**Aux** (roman + roman-**Gén** un type)-être-**St**
 (Le roman populaire est (un roman + un type de roman))

Bien que la différence sémantique entre la séquence nominale figée et la séquence nominale libre corresponde à des restrictions sur la combinatoire entre les X et les noms concrets comme roman, la nuance sémantique entre les X et des noms prédicatifs comme politique est souvent subtile. La différence de sens entre (33a) et (33b) est telle qu'on peut parler de relation de paraphrase entre eux, même si la paraphrase est approximative¹⁹.

- (33) a. *minjujeok-i-n jeongchi* (A-Sfd N)
 (démocratique-être-Sfd - politique = politique qui a une tendance démocratique)
 a'. *jeongchi-ga minjujeok-i-da*
 politique-**Nmtf** démocratique-être-**St**
 (La politique est démocratique)
 b. *minju jeongchi* (X N) (démocratique - politique = politique démocratique)
 b'. **jeongchi-ga minju-i-da*
 politique-**Nmtf** démocratique-être-**St**
 (La politique est démocratique)
 c. *minju jeongchi-neun* (jeongchi + jeongchi-ui han yuhyeong)-i-da
 politique démocratique-**Aux** (politique + politique-**Gén** un type)-être-**St**
 (La politique démocratique est (une politique + un type de politique))

Malgré la relation de paraphrase, nous observons des différences syntaxiques. Dans (33b) figé, la prédictivité de X est interdite (33b') et ce X joue un rôle sous-classificateur de N (33c). Ce n'est pas le cas pour la séquence nominale libre de (33a).

Le figement de X N est encore vérifié dans les noms figés techniques (34a) ou

¹⁸ *X N est un (N + type de N)* est un critère syntaxique de figement exposé avec le critère *N₁N₂ est (un N₂ + un type de N₂)* dans la section 2-3-3.

¹⁹ Nous considérons cette relation de paraphrase dans le chapitre 3.

les noms propres figés d'institutions officielles (34b). Les paraphrases en *Rx-jeokin N* n'ont pas de statut terminologique ou officiel. Par exemple, *gukje-jeokin yenhap* (*Rx-jeokin N*)(international - association) désigne une association d'envergure internationale ou entre plusieurs pays, alors que *gukje yenhap* (*XN*) désigne l'Organisation des Nations Unies (*ONU*).

- (34) a. *gukje-(E + *jeokin) nodongbeop* (*X N*)(international - lois régissant les conditions de travail = lois régissant les conditions de travail international)
*gukje-(E + *jeokin) baklamhoe* (*X N*)(international - exposition = exposition international)
 b. *gukje-(E + *jeokin) yenhapgun* (*X N*)(international - armée associée = Les Forces des Nations Unies ;UNF *en anglais*)
*gukje-(E + *jeokin) yenhap* (*X N*)(international - association = Organisation des Nations Unies)

Regardons les exemples du Type B (= : *Rx (*-jeokin + -han + E) N*). Les exemples de (35a) sont des groupes nominaux libres où les formes sans *-han* sont impossibles (35b), comme dans (31b).

- (35) a. *ganglyeok-ha-n* (*gyeongchal + gundae*)(*A-Sfd N*)
 (puissant- être-*Sfd* (police + armée = police + armée) puissant)
 b. *gyeongchal-i ganglyeok-ha-da*
 police-**Nmtf** puissant-être-**St**
 (La police est puissante)
 b'. **ganglyeok* (*gyeongchal + gundae*)(**X N*)
 * puissant - (police + armée)

Si nous regardons les *XN* figés du Type B, nous retrouvons les propriétés de non-prédicativité de *X* (36b'). Par ailleurs, *ganglyeok jeopchakje* (colle instantanée) est un nom de marchandise et il s'agit d'une appellation commerciale.

- (36) a. *ganglyeok-ha-n jeopchakje* (*A-Sfd N*) (être puissant-*Sfd* - colle = colle puissante)
 a' *jeopchakje-ga ganglyeokha-da*
 colle-**Nmtf** être puissant-**St**
 (La colle est puissante)
 b. *ganglyeok jeopchakje* (*X N*) (puissant - colle = colle instantanée)

b' **jeopchakje-ga* *ganglyeok-ida*
 *colle-**Nmtf** puissant-**St**
 (La colle est puissante)

Dans (36), *X* fonctionne comme un sous-classificateur de *N* :

(36) c. *ganglyeok jeopchakje-neun (jeopchakje + jeopchakje-ui han yuhyeong)-i-da*
 colle instantanée-**Aux** (colle + colle -**Gén** un type)-être-**St**
 (La colle instantanée est (une colle + un type de colle))

Regardons un autre exemple.

(37) a. *geonjeon-ha-n (olak + sasang + jeongsin) (A-Sfd N)(libre)*
 (être sain-*Sfd* (divertissement + idée + esprit =
 (chanson + divertissement + idée + esprit) sain(e))
 a'. *(olak + sasang + jeongsin)-i geonjeonha-da*
 (divertissement + idée + esprit)-**Nmtf** être sain-**St**
 (La (divertissement + idée + esprit) est saine)
 b. *(geonjeon (*olak + *sasang + *jeongsin))(X N)*
 (saine (*divertissement + *idée + *esprit))

La prédicativité de (37c) sans *-han* est aussi interdit :

c. *geonjeon gayo (XN) (figé) (sain- chanson = un genre de chanson qui n'a pas de
 tendance séditionnaire ou idéologique)*
 c'. **gayo-ga geonjeon-ida*
 *chanson-**Nmtf** sain-**St**

Même si la nuance est très subtile, le sens est un peu différent selon la structure syntaxique. Alors que *geonjeon-ha-n gayo (A-Sfd N)* de (37a) signifie n'importe quelles chansons qui n'ont pas de tendance séditionnaire ou idéologique. *geonjeon gayo (XN)* de (37d) évoque un véritable genre de chanson²⁰.

²⁰ En coréen, *gayo* et *nolae*, qui signifient « chanson », sont des synonymes. Mais *gayo* (chanson), qui est incompatible avec *-hada*, est un nom non prédicatif, et *nolae* (chanson) qui est employé avec le verbe support *-hada* est un nom prédicatif. Tandis que *gayo* est souvent utilisé comme un nom tête sous-classifié pour plusieurs types de chansons, *nolae* n'est pas un nom tête sous-classifié. Par exemple,

- (37) d. *genojeon gayo-neun* (*gayo + gayo-ui han yuhyeong*)-i-da
 chanson qui n'a pas d'idéologie-**Aux** (chanson + chanson -**Gén** un type)-être-**St**
 (Une chanson qui n'a pas d'idéologie est (une chanson + un type de chanson))

Pour le cas du Type B, la nuance subtile entre la séquence nominale figée et la séquence nominal libre est aussi sensible que celle entre *Rx-jeokin N* et *X N* (cf. (32a) - (32b)). La prédicativité de *XN* est aussi interdite au niveau syntaxique. De plus, cette structure passe le critère de figement *X N est un (N + type de N)*²¹.

Il existe des *X* du Type C (= : *Rx-(*jeokin + *han + E) N*) qui ne peuvent apparaître que dans *X N*. Ce type de *X* n'est compatible avec aucun suffixe. Pour ce dernier type de *XN*, *X* n'est jamias prédictif. Ici, nous présentons quelques exemples de ce type de *XN*, qui sont syntaxiquement tous clairement figés.

- (38) a. *gani sikdang* (*X N*)(simple - restaurant = brasserie)
 a'. **sikdang-i* *gani-ida*
 * restaurant-**Nmtf** simple-**St**
 b. *gangbak gwannyeom* (*X N*)(obsédant - idée = obsession)
 b'. **gwannyeom-i* *gangbak-ida*
 *idée-**Nmtf** obsédant-**St**
 c. *munmin jeongbu* (*X N*)(civil - gouvernement = gouvernement civil)
 c'. **jeongbu-ga* *munmin-ida*
 *gouvernement-**Nmtf** civil-**St**
 d. *bulon munseo* (*X N*)(séditieux - écrit = écrit séditieux)
 d'. **munseo-ga* *bulon-ida*
 *écrit-**Nmtf** séditieux-**St**

X fonctionne comme le sous-classificateur de *N* :

- (38) a''. *gani sikdang-eun* (*sikdang + sikdang-ui han yuhyeong*)-i-da
 restaurant simple-**Aux** (restaurant + un type de restaurant)-être-**St**

(*daejung + aechang + choesin*) *gayo* ((populaire + que l'on aime chanter + tout nouveau) – chanson = chanson (populaire + que l'on aime chanter + toute nouvelle))

* (*daejung + aechang + choesin*) *nolae* (*(populaire + que l'on aime chanter + tout nouveau) - chanson = chanson (populaire + que l'on aime chanter + toute nouvelle))

²¹ En effet, l'usage de la forme sans *-han* dans la séquence *Rx-han N* est un phénomène assez récent (No M.H (1998), p. 20).

(Une brasserie est (un restaurant + un type de restaurant))

b''. *gangbak gwannyem-eun* (*gwannyem + gwannyem-ui han yuhyeong*)-i-da

obsession -**Aux** (idée + un type d'idée)-être-**St**

(Une obsession est (une idée + un type d'idée))

c''. *munmin jeongbu-neun* (*jeongbu + jeongbu-ui han yuhyeong*)-i-da

gouvernement civil -**Aux** (gouvernement + un type de gouvernement)-être-**St**

(Un gouvernement civil est (un gouvernement + un type de gouvernement))

d''. *bulon munseo-neun* (*munseo + munseo-ui han yuhyeong*)-i-da

écrit séditieux -**Aux** (écrit + un type d'écrit)-être-**St**

(Un écrit séditieux est (un écrit + un type d'écrit))

On peut observer la même propriété pour les *X* empruntés et proprement coréens dans la structure de *XN* : la non-prédicativité de *X* et la fonction du sous-classificateur de *N*. Les *X* empruntés n'ont pas de différence avec les *X* sino-coréens dans leur relation dérivationnelle avec *-jeokida* ou *-hada*. Pour la formation des mots néologiques, un élément sino-coréen est plus facile à combiner avec un élément emprunté qu'avec un élément proprement coréen. De plus, le nombre de *X* empruntés est beaucoup plus grand que celui de *X* proprement coréens, parce que les mots d'origine étrangère perdent souvent leur autonomie dans le système du coréen sauf les noms comme *ladio* « radio », *keopi* « coffee » ou *keomputeo* « computer », etc.

Comme nous l'avions mentionné dans l'exemple (26) de la section 2-4-1, les emprunts *XN* (par exemple, *best seller*) sont des noms composés en anglais lui-même et nous empruntons directement les formes entières *XN*. Cependant, les *X* empruntés comme *balokeu* « baroque » ou *godik* « gothique » sont utilisables dans des formes dérivées *balokeu-jeok* ou *godik-jeok*. Ces *Rx* empruntés entrent par ailleurs dans des séquences nominales figées avec d'autres noms en sous-catégorisant les noms postposés. L'exemple (39) est un cas du Type A (= : *Rx(-jeokin + *-han + E) N*) où *N* est un nom sino-coréen.

(39) a. *godikjeok-i-n yangsik* (*emp sc*)(A-*Sfd N*)

(gothique - être-*Sfd* - style = style ayant des caractères de style gothique)

a'. *yangsik-i godikjeok-ida*

style-**Nmtf** être gothique-**St**

(Le style est gothique)

b. *godik yangsik* (*emp sc*)(*X N*)(gothique - style = style gothique)

b'. **yangsik-i godik-ida*

*style-**Nmtf** gothique-**St**

L'exemple (40) est un cas du Type B (= : $Rx-(\text{*jeokin} + \text{han} + E) \quad N$).

- (40) a. *legulleo-ha-n* *keopi (emp emp)(A-Sfd N)*(être normal - café = café normal)
 a'. *keopi-ga* *legulleoha-da*
 café-**Nmtf** être normal-**St**
 (Le café est normal)
 b. *legulleo* *keopi (emp-emp)(X N)*(normal - café = café américain)
 b'. **keopi-ga* *legulleo-ida*
 *café-**Nmtf** normal-**St**

L'exemple (41) est un cas du Type C (= : $Rx(\text{*jeokin} + \text{*han} + E) \quad N$).

- (41) a. *laeptap* *keomputeo (emp emp)(X N)*(portable - ordinateur = ordinateur portable ; laptop computer *en anglais*)
 b. **keomputeo-ga* *laeptap-ida*
 *ordinateur-**Nmtf** portable-**St**

Beaucoup de *XN* empruntés appartiennent au Type C, puisque leurs formes entières sont empruntées et elles ne se décomposent pas souvent.

Pour les *X* proprement coréens, il n'y a pas de structure du Type A (= : $Rx-(\text{jeokin} + \text{*han} + E) \quad N$) à cause de l'incompatibilité entre *-jeok(in)*, qui est sino-coréen, et l'élément proprement coréen. Par ailleurs, le nombre de *X* proprement coréens est petit. L'exemple (42) est un cas du Type B (= : $Rx-(\text{*jeokin} + \text{han} + E) \quad N$).

- (42) a. *kkobulang-ha-n* *geulssi (pc pc)(A-Sfd N)*
 (être courbé-*Sfd*- écriture = écriture en zigzag (ou écriture courbée par coquetterie ou orgueil))
 a'. *geulssi-ga* *kkobulang-ha-da*
 lettre-**Nmtf** courbé- être-**St**
 (L'écriture est en zigzag)
 b. *kkobulang* *geulssi (pc pc)(XN)* (courbé - lettre = lettre de l'alphabet latin)
 b'. **geulssi-ga* *kkobulang-ida*
 *lettre-**Nmtf** courbé-**St**

L'exemple (43) est un cas du Type C (= : $Rx-(*jeokin + *han + E) \quad N$).

(43) a. *sandeul balam* (XN) (souffle - vent = vent doux ; brise)

a'. **balam-i sandeul-ida*
**vent-Nmtf souffle-St*

Dans les exemples (42) et (43), X n'est pas prédicatif.

Les X empruntés et proprement coréens dans la structure de XN fonctionnent aussi le sous-classificateur de N :

(44) a. *laeptap keomputoe* (X N) (portable – ordinateur = ordinateur portable)

a'. *laeptap keomputeo-neun* (*keomputeo + keomputeo-ui han yuhyeong*)-i-da
ordinateur portable-**Aux** (ordinateur + un type d'ordinateur)-être-**St**
(Un ordinateur portable est (un ordinateur + un type d'ordinateur))

b. *kkobulang geulssi* (X N) (courbé - lettre = lettre d'alphabet latin)

b'. *kkobulang geulssi-neun* (*geulssi + geulssi-ui han yuhyeong*)-i-da
lettre alphabet latin -**Aux** (lettre + un type de lettre)-être-**St**
(Une lettre de l'alphabet latin est (une lettre + un type de lettre))

Nous pouvons conclure que les types A, B et C de XN partagent les deux propriétés syntaxiques suivantes et qu'ils sont toujours figés :

- (i) X sans aucun suffixe de dérivation n'est pas prédicatif dans XN
- (ii) X fonctionne comme sous-classificateur dans XN, et le critère *XN est (un N + un type de N)* est acceptable

3. Variations typographiques des séquences nominales figées

3-1. Différences formelles entre les composés français et coréens

Dans les sections dernières, nous avons délimité les séquences nominales figées avec la notion de figement et ses critères. Maintenant, nous allons examiner les variations typographiques des séquences nominales figées pour l'identification automatique. En vue de la constitution de dictionnaires exploitables dans des applications informatiques, l'identification automatique des séquences nominales figées exige de décrire leur structure et leurs conventions typographiques dans un dictionnaire

électronique. Selon Max Silberztein (Silberztein, 1997), le mot composé en français ou anglais est défini comme suit pour des raisons qui tiennent au traitement informatique :

Compound words are lexical units that are made up from two or more simple words (and therefore contain at least one separator.) For instance, the sequences *hard disk*, *take the bull by the horns* and *take ... into account* are compound words. Observe that the last verbal compound is not connex: it can take an arbitrary number of words between *take* and *account*. The distinction between simple and compound words is thus purely orthographic; for instance, *deltaplane* ("hang glider") is a simple word, whereas its orthographic variants *delta-plane* and *delta plane* are compound words.

Or, la définition ci-dessus ne convient pas exactement pour le nom composé du coréen, puisque l'espacement est variable selon le type des séquences nominales figées.

Nous allons commencer par mentionner les différences entre les composés français et coréens en ce qui concerne les conventions typographiques. M. Mathieu-Colas (Mathieu-Colas, 1989) présente les composés graphiques en français - à éléments séparés par une espace ou un trait d'union (exceptionnellement une apostrophe, une virgule, un point ou une barre oblique).

Dans l'usage pratique du coréen, il n'existe pas d'éléments disjoints ou reliés par un trait d'union ni par une apostrophe. Tandis qu'un trait d'union ou une apostrophe en français peuvent servir de marqueur de composition des éléments, c'est la soudure en coréen qui joue ce rôle. Les formes graphiquement soudées de la grande majorité des noms dérivés et des noms composés au sens traditionnel de la linguistique coréenne correspondent aux composés graphiques du français qui apparaissent dans la typologie de M. Mathieu-Colas. Par exemple, en coréen les formes graphiques des mots *si-eomeoni* (PF-N) « côté du mari - mère = belle-mère » ou *beol-sae* (N-N) « abeille - oiseau = oiseau-mouche »²² sont soudées, mais en français les mots ayant les mêmes sens se réalisent graphiquement avec un trait d'union (*belle-mère* ou *oiseau-mouche*). Ainsi, il n'est pas étonnant que les formes graphiquement soudées de la grande majorité des noms dérivés et des noms composés au sens traditionnel de la linguistique coréenne correspondent à des composés graphiques du français.

Deuxièmement, l'espacement entre les mots dans une séquence nominale est très irrégulier, comme nous l'avons vu dans le chapitre 1. Pour la grande majorité des séquences de noms, en principe, on met un espace, mais il est acceptable de ne pas

²² *beol sae* (N N) sous forme non soudée signifie « une abeille et un oiseau ».

espacer. Par exemple, une même séquence nominale figée 공중 전화 카드 *gongjung jeonhwa kadeu* ($N\ N\ N = N$)(*publique - téléphone - carte = carte pour le téléphone public*) peut se réaliser graphiquement de quatre façons possibles :

- (45) 공중 전화 카드 ($N\ \ \ N\ \ \ N$)
공중 전화카드 ($N\ \ \ NN$)
공중전화 카드 ($NN\ \ \ N$)
공중전화카드 (NNN)

3-2. Variations typographiques

Dans cette section, nous examinons les désaccords entre les dictionnaires éditoriaux et entre les dictionnaires éditoriaux et les textes.

3-2-1. Désaccords entre les dictionnaires éditoriaux

Par une comparaison simple de l'espacement dans cinq dictionnaires éditoriaux parus depuis 1990, nous découvrons des désaccords nombreux. Extrayons seulement quelques exemples dans la Figure 2.

Categories Gram.	Dic 1 (91)	Dic 2 (97)	Dic 3 (98)	Dic 4 (99)	Dic 5 (95)	Traduction
$N-N$	겨울방학	겨울 방학	겨울 방학	겨울^방학	겨울 방학	vacances d'hiver
$N-N$	국회의원	국회 의원	국회 의원	국회의원	국회 의원	membre de l'Assemblée National
$X-N$	꼬부랑글자	꼬부랑 글자	꼬부랑글자	꼬부랑글자	꼬부랑 글자	lettres de l'alphabet latin
$N-N$	눈언저리	눈언저리	눈언저리	눈언저리	눈 언저리	contour des yeux
$X-N$	원생동물	원생 동물	-	원생동물	원생동물	protozoaire
$X-N$	강장동물	강장 동물	-	강장동물	강장동물	cœlentérés
$N-N$	먹이사슬	먹이 사슬	먹이 사슬	먹이^사슬	먹이사슬	chaîne alimentaire
$N-N$	사돈댁**	사돈댁	사돈댁	사돈댁	사돈 댁	belle -famille
$N-N$	양은남비**	양은 냄비	양은 냄비***	양은 냄비***	양은냄비	pot en argent blanc
$V-Sfd-N$	죽을힘	죽을힘	죽을 힘	죽을힘	죽을 힘	dernier effort
$Det-N$	허튼수작	허튼 수작	허튼수작	허튼수작	허튼수작	baratin

Figure 2. Désaccords entre dictionnaires éditoriaux²³

²³ Dans cette figure, ** indique des variantes orthographiques et *** un exemple du dictionnaire (non pas une entrée du dictionnaire).

En général, les exemples énumérés dans la figure 2 appartiennent aux noms composés dans la linguistique traditionnelle. Il est clair que les conventions typographiques ne sont pas représentées de manière unanime par les dictionnaires²⁴.

3-2-2. Désaccords entre les dictionnaires et les textes

Nous découvrons également des cas où les formes typographiques de quelques noms composés varient selon les textes, même lorsqu'il y a un accord dans la convention typographique entre les dictionnaires. Nous avons utilisé le corpus de KAIST (<http://morphkaist.ac.kr/kcp>), qui contient 70 millions d'unités typographiques simples²⁵. Le résultat est le suivant :

47²⁶: 지어 선 **소나무** , 잣나무 , 동백나무 , **혜나무** , 느티나무 가지
 44: 부터 분주했다 . \ **소나무** , 동백나무 , **혜나무** , 자작나무 등에
 82: 늘을 향해 직립해 있는 붉은 **소나무** 숲과 울창한 마로니에 숲으
 89: 을 뜯고 있던 암소도 **아카시아 나무** 그늘 밑으로 피해 있었다 .
 68: 무며 상수리나무며 박달나무며 **아카시아나무**며가 뒤섞인 숲이
 24: 른 뜻을 말했다 . \ "내가 **보리수 나무** 그늘에서 깨달음을 이룬 지
 89: 면서 말했다 . \ " 붓다는 **보리수나무** 밑에서 득도를 하셨습니다
 87: 수목과 바위가 보였다 . \ **보리수나무**들이 서 있는 곳에 세 개의
 110: 은 다시 눈을 뜨고 멀리 **야자수 나무** 밑에 서 있는 웨곤으로 시
 57: 붙어왔다 . \ 바나나와 **야자수 나무**가 모래사장 한 옆에 가득
 57: 가본 자카르타 동북 해안 **야자수나무** 숲이 있는 그 위안소는 커
 127: 경상도지방에서는 꿈에 **도토리 나무**를 보면 행운이 따른다고 믿
 127: 스리고 하혈을 멈추게 한다 . \ **도토리나무**의 껍질을 달여먹으면
 67: 있었고 , 주위에는 상수나무 , **도토리나무** , 밤나무 , 떡갈나무
 127: 날 중국의 서하지방의 여인들은 **구기자 나무**의 열매 , 잎 , 뿌리
 127: 차 (茶) 를 만들어 마신다 . \ **구기자나무**의 열매나 잎을 재료
 69: 저 세상도 다 버려라 . \ **무화과 나무**에 꽃이 없듯이 이 세상에
 118: 방 저쪽 끝에는 아직 젖어 있는 **무화과나무**의 가지가 창을 거쳐

Figure 3. Corpus de KAIST

Toutes les séquences nominales en gras (soulignées ou dans des boîtes) sont des noms

²⁴ L'exemple *haneul nala* (NN) « ciel - pays = paradis » montre un désaccord dans l'espacement entre l'entrée *haneulnala* et l'exemple *haneul nala* dans Dic 3.

²⁵ L'unité typographique simple signifie le nom-*Postp*, le verbe-*Sflx*, l'adjectif-*Sflx*, etc.

²⁶ Ce numéro indique le numéro du fichier.

composés au sens traditionnel. Par exemple :

- (46) a. *dotoli-namu* « gland » « arbre » = « chêne »
 yajasu-namu « noix de coco » « arbre » = « cocotier »
 b. *so-namu* « pin » « arbre » = « pin »
 dae-namu « bambou » « arbre » = « bambou »

Les exemples (46a) et (46b) sont imprimés sans blanc dans tous les dictionnaires commerciaux. Mais les séquences (46a) apparaissent également avec des blancs dans des textes comme dans la figure 3, et le dictionnaire électronique ne devrait pas négliger cette possibilité. Nous indiquons par « ^ » la position de l'espacement facultatif observé dans des textes.

- (47) a. *dotoli ^ namu* (3+2) « gland » « arbre » = « chêne »
 yajasu ^ namu (3+2) « noix de coco » « arbre » = « cocotier »
 b. *so-namu* (1+2) « pin » « arbre » = « pin »
 dae-namu (1+2) « bambou » « arbre » = « bambou »

Il nous paraît que le nombre de syllabes joue un rôle : quand un mot est figé (ou composé) avec un mot monosyllabique, il ne contient en général aucune espace comme les exemples dans (47b). Nous vérifions ce fait par les exemples toujours soudés :

- (48) *machum-beop* (2+1)²⁷(*sc-sc*) « orthographe » « règle » = « règle d'orthographe »
 gyotong-beop (2+1)(*sc-sc*) « circulation » « règle » = « règle de circulation »
 kkot-galu (1+2)(*pc-pc*) « fleur » « poudre » = « pollen »

En cas de discordance entre l'usage et les grammairiens, on peut analyser ces formes soit comme des erreurs de l'utilisateur, soit comme des variantes acceptables. Cependant, pour identifier les formes qui contiennent un espacement et les formes qui sont soudées à d'autres formes, on a besoin d'un dictionnaire qui précise les possibilités d'espacement.

3-3. Classification des séquences nominales figées en fonction des conventions typographiques

²⁷ Le chiffre indique le nombre de syllabes.

Nous avons examiné la notion linguistique de figement pour délimiter les séquences nominales figées, et les variations typographiques dans les textes. Maintenant, nous allons distinguer les séquences nominales figées selon les conventions typographiques. Voici des exemples de séquences nominales figées toujours soudées.

- (49) a. *si-eomeoni* (Pf-N) « côté du mari » « mère » = « belle-mère »
yengu-ja (N-Sf) « recherche » « personne » = « chercheur »
jeongchi-ga (N-Sf) « politique » « personne » = « politicien »
mu-uitak-ja (Pf-N-Sf) « néant » « compte sur » « personne » = « personne qui n'a personne sur qui compter »
- b. *kkot-galu* (N-N) (1+2) « fleur » « poudre » = « pollen »
dukkeobi-jip (N-N) (3+1) « crapaud » « maison » = « boîte à fusibles »
gyotong-beop (N-N) (2+1) « circulation » « règle » = « règle de circulation »
geom-beoseot (Ra-N) (1+2) « noir » « champignon » = « taches noires sur la peau des vieillards »
mat-balam (Rv-N) (1+2) « recevoir » « vent » = « vent contraire ou de face »
swim-pyo (Vsfn-N) (1+1) « pause » « signe » = « pause »

Les exemples (49a) sont des noms dérivés et (49b) des noms composés au sens traditionnel. Les noms dérivés sont toujours soudés. La majorité des noms composés au sens traditionnel²⁸ sont soudés : tandis que les noms composés entre un mot et un mot monosyllabique sont en général soudés, les noms composés d'éléments à plusieurs syllabes peuvent être des séquences nominales figées à espacement facultatif, même si ce sont des noms composés au sens traditionnel (cf. *dotoli ^ namu* (3+2) « gland - arbre = chêne » dans la section 3-2).

Regardons des exemples de séquences nominales figées à espacement facultatif.

- (50) a. *gukhoe ^ uijang* (N^N) (2+2) « l'Assemblée nationale » « président »
= « député qui représente l'Assemblée nationale »
gajok ^ gyehoek (N^N) (2+2) « famille » « projet » = « contrôle des naissances »
chuli ^ soseil (N^N) (2+2) « déduction » « roman » = « roman policier »
dotoli ^ namu (N^N) (2+2) « gland » « arbre » = « chêne »

²⁸ cf. section 1 dans le chapitre 1.

- b. *gongsang* ^ *gwahak* ^ *yeonghwa* ($N^N N$) (2+2+2)
 « imagination » « science » « film » = « film de science-fiction »
gongjung ^ *jeonhwa* ^ *kadeu* ($N^N N$) (2+2+2)
 « public » « téléphone » « carte » = « carte pour le téléphone public »
- c. *munmin* ^ *jeongbu* (X^N) « civil » « gouvernement » = « gouvernement civil »
tongsok ^ *soseol* (X^N) « populaire » « roman » = « roman populaire »
munmin ^ *jeongbu* (X^N) « civil » « gouvernement » = « gouvernement civil »
bulon ^ *munseo* (X^N) « séditieux » « écrit » = « écrit séditieux »
beseuteu ^ *seleo* (X^X) « meilleur » « vendeur » = « auteur à succès »
bitamin ^ *di* (N^X) « vitamine » « D » = « vitamine D »
- d. *dolyeon* ^ *byeoni* (B^N) « soudain » « mutation » = « mutation sans préavis »
bolok ^ *geoul* (B^N) « convexe » « miroir » = « miroir convexe »

Les exemples (50a) - (50d) sont des noms figés à espacement facultatif et les possibilités d'espacement des séquences de deux noms (50a) sont au nombre de deux et celles des séquences de trois noms (50b) de quatre. Parmi 45000 séquences nominales figées à espacement facultatif, 15 % contiennent au moins un X , c'est-à-dire XN , XX , NX , XNN , etc. et on peut insérer un blanc entre X et N , N et X ou X et X . Assez rarement on trouve des entrées à espacement facultatif de la structure BN où B indique l'initiale de l'adverbe *busa* en coréen (50d).

Les exemples (51a) et (51b) sont des séquences nominales figées à espacement obligatoire.

- (51) a. *saeppalga-n geojinmal* ($A-Sfd\ N$) « être rouge foncé-*Sfd* » « mensonge »
 = « gros mensonge »
- b. *geulim-ui tteok* ($N-Gén\ N$) « tableau-*Gén* » « gâteau de riz » = « gâteau de riz du tableau ; objet que l'on peut utiliser ou avoir »
- bingsan-ui ilgak* ($N-Gén\ N$) « iceberg-*Gén* » « coin » = « la partie émergée de l'iceberg »

Lorsqu'un adjectif ou un verbe qualifie un nom comme (51a), il est ordinairement suivi d'un suffixe déterminatif et l'espacement entre le modifieur et le nom est obligatoire (figé ou libre) sauf dans le cas des noms composés au sens traditionnel²⁹. Dans (51b),

²⁹ L'insertion d'un blanc entre $A-Sfd$ et N dans les noms composés au sens traditionnel de la forme $A-Sfd\ N$ fait disparaître le sens figé. Comparons la différence du sens selon l'espacement :

Considérons les exemples dont les structures internes sont plus compliquées et dont l'espacement est plus irrégulier.

- Ce sont des séquences nominales figées qui contiennent au moins un blanc. Dans les dictionnaires éditoriaux, on met un blanc entre *juk* (N) et *meokgi* (N), *pulchil* (N) et *hagi* (N), *ko* (N) et *mut-eun* (V-Sfd), et *kkul* (N) et *meok-eun* (V-Sfd). Mais dans l'usage, après l'effacement d'une postposition nominale, on a une tendance à coller l'élément précédant la postposition nominale et l'élément suivant (par exemple, *juk* (N)-effacement de *-eul* (Acc)-*meokgi* (N) => *juk-meokgi*). L'insertion des postpositions nominales fait disparaître le sens figé.

Mais si les adjectifs ou les verbes qualifient directement les noms sans suffixe déterminatif, ce sont des noms composés à la composition déviante. Rappelons les exemples dans le chapitre 1.

65

- (53) *sik-eun juk-eul meokgi* (V-Sfd N-Acc N)
 « se refroidir-Sfd » « bouillie de riz » « manger-Sfn » =
 « action de manger la bouillie de riz refroidi »
- ip-e pulchil-eul hagi* (N-postp.e N-Acc N)
 « à la bouche » « action d'étendre de la colle » « faire-Sfn » = « action d'étendre de la colle à la bouche »
- ko-ga mut-eun don* (N-Nmtf V-Sfd N)
 « morve » « s'attacher à-Sfd » « argent » = « argent à laquelle la morve s'attache »
- kkul-eul meok-eun beongeoli* (N-Acc V-Sfd N)
 « miel » « manger-Sfd » « sourd-muet » = « être muet qui a mangé le miel »

Enfin, selon les critères de figement et les conventions typographiques, nous définissons des noms compacts et des noms figés. Les *noms compacts* sont des noms simples, des noms dérivés et des séquences nominales figées toujours soudées, alors que les *noms figés* sont des séquences nominales figées à espacement facultatif ou à espacement obligatoire. Les séquences nominales figées sont donc classées en trois catégories selon des conventions typographiques :

(i) *Noms compacts (NC)*

Si une séquence nominale figée ne contient aucun espacement obligatoire ni aucun espacement facultatif, alors c'est un nom compact.

- Tous les noms simples sans dérivation ni composition : *eomeoni* (N) « mère »
- Tous les noms dérivés avec des affixes :
jeongchi-ga (N-Sf) « politique » « personne » = « politicien »
- Les noms composés toujours soudés :
kkot-galu (NN) « fleur » « poudre » = « pollen »

(ii) *Noms figés à espacement facultatif (NFF)*

Si une séquence nominale figée contient au moins un espacement facultatif sans aucun espacement obligatoire, alors c'est un nom figé à espacement facultatif.

- *pali ^ moksum* (NN) « mouche » « vie » = « vie de peu de valeur »

(iii) *Noms figés à espacement obligatoire (NFO)*

Si une séquence nominale figée contient au moins un espacement obligatoire, alors c'est un nom figé à espacement obligatoire.

- *bingsan-ui ilgak* (N-Gén N)
« iceberg-Gén » « coin » = « la partie émergée de l'iceberg »

4. Conclusion

Voici la structure des séquences nominales figées dans un dictionnaire électronique :

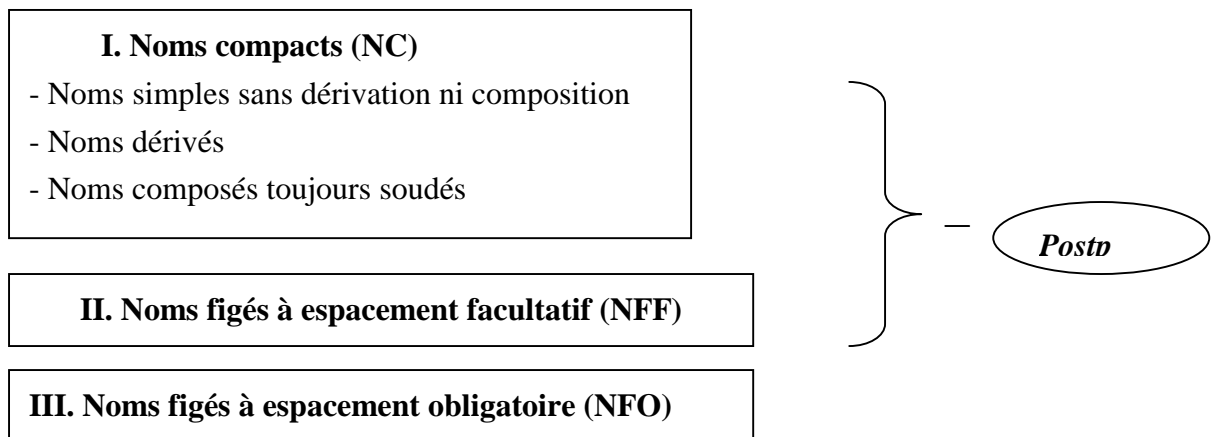


Figure 4. Structure des séquences nominales figées dans un dictionnaire électronique

Tous ces trois types de séquences nominales figées se combinent avec les séquences de postpositions nominales de la même façon, à savoir les séquences de postpositions nominales sont toujours accolées à la dernière syllabe des séquences nominales figées³⁰.

³⁰ cf. Pour la combinaison entre les séquences nominales figées et les séquences de postpositions nominales, voir le chapitre 3 et 5.

Chapitre 3. Description des séquences nominales figées dans un dictionnaire électronique

Dans ce chapitre, nous discutons les informations à inclure dans un dictionnaire électronique. La description est focalisée sur les noms figés à espacement facultatif (*NFF*)¹, mais nous présentons des perspectives sur la description des autres types de séquences nominales figées. Pour construire le dictionnaire des séquences nominales figées, nous proposons sept informations fondamentales :

- l'espacement
- les informations flexionnelles
- l'origine
- la catégorie grammaticale
- les informations sémantiques
- les informations morpho-syntaxiques
- les informations syntaxiques

Nous allons voir pourquoi nous avons choisi ces informations et comment nous les incluons dans un dictionnaire électronique.

1. Espacement

Dans le chapitre 2, nous avons classé les séquences nominales figées en fonction des conventions typographiques. Pour identifier précisément au cours du traitement automatique les variantes des séquences nominales liées à l'espacement, il est indispensable d'inclure les informations sur l'espacement dans un dictionnaire électronique des séquences nominales figées.

Dans notre dictionnaire, nous indiquons par « ^ » la position de l'espacement facultatif et par un blanc la position de l'espacement obligatoire observé dans des textes et des dictionnaires éditoriaux :

- (1) a. *jeongchi-ga* (N) « politique » « personne » = « politicien »
b. *kkot-galu* (N) « fleur » « poudre » = « pollen »
(2) a. *gukhoe ^ uijang* (N+NN) « l'Assemblée nationale » « président »

¹ Pour le moment, nous avons décrit les 45000 *NFF* avec ces sept informations.

- = « député qui représente l'Assemblée nationale »
- b. *gongsang* ^ *gwahak* ^ *yeonghwa* (N+NNN)
 « imagination » « science » « film » = « film de science-fiction »
- c. *munmin* ^ *jeongbu* (N+XN) « civil » « gouvernement » = « gouvernement civil »
- d. *dolyeon* ^ *byeoni* (N+BN) « soudain » « mutation » = « mutation sans préavis »
- (3) a. *saepalga-n* *geojinmal* (N+A-nN) « être rouge foncé-Sfd » « mensonge » =
 « gros mensonge »
- b. *geulim-ui* *tteok* (N+N-uiN) « tableau-Gén » « gâteau de riz » = « gâteau de riz
 du tableau ; objet que l'on ne peut avoir »
- (4) a. *sik-eun* *juk* ^ *meokgi* (N+V-eunNN) « se refroidir-Sfd » « bouillie de riz » « manger-
Sfn » = « travail très facile à faire »
- b. *ip-e* *pulchil* ^ *hagi* (N+N-eNN) « à la bouche » « action d'étendre de la colle »
 « faire-Sfn » = « vie dans la misère »

Dans l'écriture du coréen, le tiret n'apparaît pas et le tiret dans les exemples coréens ci-dessus indique la frontière des morphèmes. Les (1) sont des noms compacts (NC) : (1a) est un nom compact dérivé et (1b) un nom compact composé. Les exemples (2a) - (2d) sont des noms figés à espacement facultatif (NFF). Les possibilités d'espacement pour une séquence de deux noms (2a) sont au nombre de deux et pour trois noms (2b) de quatre. Les exemples (3a) et (3b) sont des noms figés à espacement obligatoire. Les structures internes des exemples (4) sont plus compliquées et les espacements sont plus irréguliers. Dans les exemples (2) - (4), le symbole « + » lie la catégorie grammaticale (N) et leurs sous-catégories (NN, N-uiN, etc.).

2. Informations flexionnelles

2-1. La combinatoire entre les noms et les postpositions nominales²

Pour la reconnaissance automatique, le dictionnaire électronique des séquences nominales figées doit inclure les informations flexionnelles puisque les formes fléchies apparaissent dans les textes. En coréen, tous les substantifs prennent, en principe, une ou une combinaison de plusieurs postpositions nominales. Les postpositions nominales sont comparables aux prépositions en français :

² cf. Nous traitons la classification et la combinatoire des postpositions nominales dans le chapitre 4. Ici, nous ne mentionnons que la combinatoire entre les séquences nominales figées et les postpositions nominales.

- (5) a. *chingu-ege* « ami-*Postp.e* » = « à l'ami »
 b. *chingu-ege-neun* « ami-*Postp.e-Aux* » = « au moins à l'ami »
 c. *chingu-ege-man-eun* « ami-*Postp.e-Aux* » = « au moins seulement à l'ami »

Il est donc indispensable de construire le dictionnaire électronique des formes fléchies des séquences nominales figées. En souci d'efficacité, nous proposons de construire deux dictionnaires distincts : un pour les séquences nominales figées, représenté par une liste ou des graphes, et un autre pour les séquences de postpositions nominales représenté par des graphes. Ces deux dictionnaires peuvent être fusionnés en un seul à l'aide des codes flexionnels attachés à chaque séquence nominale figée et du module de flexion (« Inflection ») d'INTEX. Pour chaque séquence nominale figée, il faut décrire son code flexionnel dans un dictionnaire électronique. Or, pour décrire le code flexionnel, il faut considérer les restrictions combinatoires entre les noms et les postpositions nominales. Nous observons deux types de restrictions combinatoires suivant qu'elles sont basées sur le contexte phonologique ou sur les traits sémantiques.

Regardons les contraintes phonologiques. L'emploi de 38 des postpositions nominales dépend de conditions phonologiques lors de la concaténation avec d'autres éléments, et les autres postpositions nominales n'en dépendent pas. Par exemple, la postposition auxiliaire <-neun>³ possède les deux formes *-neun* / *-eun* selon le contexte gauche.

-neun / V_____ (V : voyelle)
-eun / C_____ (C : consonne)

- (6) a. *chingu-neun* « ami-*Aux* » = « au moins l'ami »
 b. *siktak-eun* « table-*Aux* » = « au moins la table »

Nous avons choisi arbitrairement la forme (ou une des formes) après voyelle comme forme canonique des postpositions nominales.

Dans le cas des postpositions nominales *-lo* « avec, à, etc. », *-loseo* « comme, de, etc. », *-losseo* « par, avec, etc. », nous constatons qu'il existe un contexte gauche exceptionnel avec la consonne *l*. Les séquences terminées par une consonne sont généralement combinables avec les formes *-eulo*, *-euloseo*, *-eulosseo*. Ce sont des variantes phonologiques de <-lo>, <-loseo> et <-losseo>. Cependant, si une séquence

³ « < > » indique la forme canonique.

est terminée par la consonne *l*, elle n'est combinable qu'avec les formes *-lo*, *-loseo*, *-losseo* comme les séquences terminées par une voyelle.

-eulo, *-euloseo*, *-eulosseo* / C_____ (C : consonnes autres que *l*)
-lo, *-loseo*, *-losseo* / V_____
-lo, *-loseo*, *-losseo* / l_____

Prenons des exemples:

(7) a. *geomchal chongjang-euloseo*

« procureur général-*Postp.loseo* » = « comme le Procureur général »

b. *gajeong jubu-deul-loseo*

« femme au foyer-*Pl-Postp.loseo* » = « comme des femmes au foyer »

c. *gajeong jubu-loseo*

« femme au foyer-*Postp.loseo* » = « comme une femme au foyer »

Outre les contraintes phonologiques, d'autres contraintes sur la combinaison avec des postpositions nominales sont liées aux traits sémantiques. Pour le moment, notre dictionnaire contient 6 traits sémantiques : [+Hum], [+Plt], [+Anm], [+Inanim], [+CollHum] et [+InstHum] (cf. section 5 pour la classification et les critères associés aux traits sémantiques). Par exemple, les postpositions *-ege*, *-hante* et *-e* sont les trois variantes de la forme canonique <-e>. Les formes *-ege* et *-hante* sont des variantes stylistiques : *-ege* pour la langue écrite *-hante* pour la langue parlée. La compatibilité des noms avec les postpositions *-ege* et *-e* nous permet d'assigner des traits sémantiques à certains noms.

(8) a. *chingu-ege*

[+Hum] « ami-*Postp.e* » = « à l'ami »

b. *siktak-e*

[+Inanim] « table-*Postp.e* » = « à la table »

c. *minjok-ege*

[+CollHum] « peuple-*Postp.e* » = « à un peuple »

d. *jeongbu-e*

[+InstHum] « gouvernement-*Postp.e* » = « au gouvernement »

La postposition nominale *-eseo* ne marque le nominatif qu'avec les noms humains collectifs institutionnels.

(8) d. *jeongbu-eseo*

[+InstHum] « gouvernement-*Nmtf* »

d'. *jip-eseo*

[+Inanim] « maison-*Postp.eseo* » = « dans la maison »

Les exemples suivants illustrent deux contraintes combinatoires :

- (8) a'. *chingu-ege-neun* [+Hum] « ami-*Postp.e-Aux* » = « au moins à l'ami »
 b'. *siktak-e-neun* [+Inanim] « table-*Postp.e-Aux* » =
 « au moins à la table »
 a''. *chingu-ege-man-eun* [+Hum] « ami-*Postp.e-Aux-Aux* » =
 « au moins seulement à l'ami »
 b''. *siktak-e-man-eun* [+Inanim] « table-*Postp.e-Aux-Aux* » =
 « au moins seulement à la table »

La postposition auxiliaire <-neun> se réalise -neun après la voyelle *e* dans (8a') et (8b'), mais -eun après la consonne -*n* dans (8a'') et (8b'').

2-2. Classification des séquences nominales figées selon les contraintes combinatoires

A partir des deux types de contraintes sur la combinaison avec les postpositions nominales, nous classons les séquences nominales figées en neuf catégories selon les contraintes combinatoires phonologiques et sémantiques :

- N11 : Terminaison avec consonne sauf *l*, (+Hum, +Anm, +CollHum)
 (3400 séquences de postpositions nominales)
- N12 : Terminaison avec consonne sauf *l*, (+InstHum)(1650)
- N13 : Terminaison avec consonne sauf *l*, (+Plt, +Inanim) (1900)
- N21 : Terminaison avec consonne *l*, (+Hum, +Anm, +CollHum)(3400)
- N22 : Terminaison avec consonne *l*, (+InstHum)(1650)
- N23 : Terminaison avec consonne *l*, (+Plt, +Inanim) (1900)
- N31 : Terminaison avec voyelle, (+Hum, +Anm, +CollHum)(3400)
- N32 : Terminaison avec voyelle, (+InstHum)(1650)
- N33 : Terminaison avec voyelle, (+Plt, +Inanim) (1900)

Pour chaque classe, le premier chiffre (1-3) indique la condition phonologique et le deuxième chiffre (1-3) les traits sémantiques. Le chiffre entre parenthèses est le nombre de séquences de postpositions possibles pour un nom. Par exemple, la classe N11 couvre toutes les séquences nominales figées qui satisfont à la fois les deux conditions :

qui terminent par consonne sauf *l* et dont les traits sémantiques sont des noms humains, des noms humains collectifs ou des noms d'animaux. Le nombre de séquences de postpositions pour un nom de la classe N11 est 3400. Le nombre moyen de séquences de postpositions pour un nom est donc environ 2300. Voici quelques exemples d'entrées codifiées :

- (9) a. *geomcha l[^] chongjang*, N11+NN+Hum « Procureur général »
 b. *gukchak ^ eunhang*, N12+NN+InstHum « banque nationale »
 c. *gyeoul ^ banghak*, N13+NN+Inanim « vacances d'hiver »
 d. *gunjip ^ saenghwal*, N23+NN+ Inanim « vie de la troupe »
 e. *gajeong ^ jubu*, N31+NN+Hum « femme au foyer »
 f. *gyeongchal ^ gukga*, N32+NN+InstHum « pays basé sur la force de la police »
 g. *gunae ^ jeonhwa*, N33+NN+ Inanim « téléphone dans l'intérieur du bâtiment »

A partir de codes flexionnels comme N11 ou N33 dans un dictionnaire électronique et de la fonctionnalité de flexion du menu DELA d'INTEX (« Inflection »), nous pouvons engendrer les formes fléchies des séquences nominales figées (pour les détails, voir le chapitre 5). Cette fonctionnalité permet la flexion automatique en appelant des transducteurs finis de postpositions nominales qui font un dictionnaire électronique des séquences de postpositions nominales. Le nom du transducteur est exactement le code flexionnel associé à chaque entrée de séquence nominale figée.

3. Origine

Le lexique coréen se subdivise en quatre parties selon l'origine : le vocabulaire proprement coréen, le vocabulaire sino-coréen, le vocabulaire emprunté à d'autres langues que le chinois comme l'anglais, le français, etc. (emp), et les mots constitués d'éléments mélangés (Mel). Dans notre dictionnaire, nous utilisons l'initiale de la traduction anglaise comme codage de l'origine : sk (*sino-Korean*), pk (*proper Korean*), bor (*borrowed*), Mix (*Mixed*).

- (10) a. *gajeong jubu* (sk_sk)(NN) « famille » « ménagère » = « femme au foyer »
 b. *daechu namu* (pk_pk)(NN) « jujube » « arbre » = « jujubier »
 c. *gaseu boilleo* (bor_bor)(NN) « gaz » « chaudière » = « chaudière à gaz »
 d. *gaen taeksi* (sk_bor)(NN) « individuel » « taxi » = « taxi dont le propriétaire est le chauffeur lui-même, non

- pas un syndicat de taxis »
- gyeoul banghak* (*pk_sk*)(*NN*) « hiver » « vacances » = « vacances d'hiver ;
vacances de Noël et du Jour de l'An »
- e. *chong-al taeksi* (*Mix(sk_pk)_bor*) (*NN*) « balle » « taxi » = « taxi qui roule très
vite comme un balle »
- cho-beol gui* (*Mix(sk_pk)_pk*) (*NN*) « action première » « cuisson » =
« cuisson des céramiques »

Pour les 45000 *NFF*, nous comptons 80% de **sk**, 6% de **pk**, 13% de **bor**, et 1% de **Mix**⁴. Les vocabulaires sino-coréen et emprunté constituent une grande partie du lexique du coréen. Les *NFF* du type *pk_pk* comme (12b) sont rares puisqu'en général, les noms composés proprement coréen sont compacts.

L'origine du vocabulaire coréen peut permettre d'identifier le sens exact du mot. Par exemple, la forme *susul* et la forme *gomu* peuvent avoir respectivement trois sens et deux sens différents en fonction de l'origine :

- (11) a. *su-sul* 수술 *pk* (*N*) « étamine de la plante »
a'. *su-sul* (繡술) *Mix (sk_pk)* (*N*) « frange »
a''. *su-sul* (手術) *sk* (*N*) « main » « technique ou art » = « opération »
b. *gomu* 고무 *pk* (*N*) « caoutchouc »
b.' *gomu* (鼓舞) *sk* (*N*) « encouragement »

Les origines *pk*, *Mix(sk-pk)* et *sk* des trois entrées *susul* permettent de séparer trois sens différents, et *pk* et *sk* des deux entrées *gomu* deux sens différents.

La description sur l'origine n'est pas toujours suffisante pour séparer les sens. Par exemple, *sajeon* peut avoir plusieurs sens selon les caractères chinois⁵. Ce fait

⁴ Les statistiques seront différentes pour les noms compacts.

⁵ *Yonsei Hangukeo Sajeon* (Yonsei Dictionnaire du Coréen)(1998) de Centre de recherches sur le coréen de l'Université Yonsei.

cf. Dans *Pyojun Gugkeo Daesajeon* (Grand Dictionnaire Standard du Coréen) (1999) de l'Académie Nationale de la Langue Coréen, les 22 sens pour *sajeon* sont décrits selon les caractères chinois :

sajeon 1(史前) (*N*) préhistoire
sajeon 2(史傳) (*N*) histoire et biographie
sajeon 3(寺田) (*N*) champ appartenant à un temple
sajeon 4(寺典) (*N*) livre éminent
sajeon 5(死前) (*N*) avant la mort
sajeon 6(死戰) (*N*) combat acharné
sajeon 7(私田) (*N*) champ appartenant à un particulier
sajeon 8(私電) (*N*) télégramme personnel
sajeon 9(私戰) (*N*) terrorisme

demande de décrire les caractères chinois dans le dictionnaire électronique⁶ :

(12) *sajeon* 1(辭典) (N) dictionnaire

sajeon 2(事典) (N) encyclopédie

sajeon 3(事前) (N) avant le travail (ou le fait) = à l'avance

sajeon 4(私田) (N) champ appartenant à un particulier

Si cette forme *sajeon* est écrite en caractères chinois dans les textes, l'ambiguïté peut être levée.

En effet, on rencontre parfois des caractères chinois pour les mots **sk**. Quand les caractères chinois sont utilisés, ils accompagnent (type A) ou remplacent (type B) les mots écrits dans l'alphabet coréen⁷ :

Type A⁸.

또한 불교의 무아(無我) 사상도 일체의 자아를 인정하지 말라는 뜻이라기보다 진자아(眞自我)가 아닌 가아(假我)에 망령되어 집착하지 말라는 뜻이다. 그러므로 무아 사상은 사회 의식을 부정하기보다는 아집과 망집에 사로잡힌 인간 관계를 제거해야 한다는 뜻으로 해석할 수 있다. 다시 말해서, 이러한 무아 사상은 관계를 갖지 않은 실재는 이 세상에 하나도 있을 수 없다는 연기 사상이, 투쟁을 필요로 하는 사회의 모순을 단순히 무명(無明)의 반영으로 보려는 소극적인 자세를 제거시킴으로써, 사회에 대한 보다 적극적인 자세를 가져야 한다는 사상으로 볼 수 있다.

sajeon 10(私錢) (N) fausse monnaie

sajeon 11(私轉) (N) rotation sur soi

sajeon 12(事典) (N) encyclopédie

sajeon 13(事前) (N) avant le travail (ou le fait) = à l'avance

sajeon 14(祀典) (N) offrande aux ancêtres

sajeon 15(沙田/砂田) (N) champ de sable

sajeon 16(師傳) (N) action de recevoir des instructions par d'un enseignant

sajeon 17(梭田) (N) champ en forme de navette de tisserand

sajeon 18(赦典) (N) amnistie le jour de la fête nationale de la part du roi

sajeon 19(肆塵) (N) magasin

sajeon 20(賜田) (N) champ que le roi en récompense (époque Lee)

sajeon 21(謝電) (N) télégramme de remerciements

sajeon 22(辭典) (N) dictionnaire

⁶ Dans INTEX, nous pouvons afficher 2350 caractères coréens et 4888 caractères chinois utilisés dans la vie quotidienne en les définissant dans le fichier **Alphabet**. Pour les détails, voir l'annexe 1. Définition de l'alphabet du coréen dans INTEX sous Windows 98 ou 2000 en coréen.

⁷ Les exemples des types A et B qui sont tirés dans le corpus de *Sejong Projet 21* de l'Académie Nationale de la Langue Coréen contenant 10 millions d'unités typographiques simples (<http://www.sejong.or.kr> ou <http://www.korean.go.kr>).

⁸ cf. Kim Heung-Kyu et Yoon Ku-Byeong (1993).

Type B⁹.

(...)意思疏通을 위해 적어도 하나 이상의 言語는 반드시 배우고 사용하지 않을 수 없다. 일상의 주고 받는 말, 과거의 역사적 사실, 그리고 미래에 대한 예상도 언어로 표현되지 않는 것이 없고, 이러한 表現으로써 話者和 聽者의 意思疏通이 가능해지고 또한 이 言語를 통해서만 시대를 달리하는 知識의 전달과 文化의 계승 발전도 가능하게 된다.

Le type A est beaucoup plus fréquent que le type B. Mais même pour traiter un texte du type B, le dictionnaire doit inclure la forme en alphabet coréen pour pouvoir normaliser les caractères chinois en formes coréennes.

4. Catégorie grammaticale des éléments

Dans un dictionnaire électronique, nous codifions les catégories grammaticales comme suit : *N* (Nom), *A* (Adjectif), *D* (Déterminant), *B* (adverbe), *V* (Verbe), etc. Nous codons l’adverbe *B*, pour éviter l’ambiguïté avec *A* (Adjectif), *D* (Déterminant), et *V* (Verbe). *B* est l’initiale de *busa* « adverbe » en coréen.

La catégorie grammaticale permet de spécifier certaines ambiguïtés lexicales. La forme *gajang* est lexicalement ambiguë :

- (13) a. *gajang* (가장)(*pk*) ***B*** « le plus, extrêmement »
- b. *gajang* (家長)(*sk*) ***N*** « chef de famille »
- c. *gajang* (假裝)(*sk*) ***N*** « déguisement »
- d. *gajang-hada* (假裝-hada)(*sk*) ***V*** « se déguiser »

Ces exemples illustrent deux types d’ambiguïté. Le premier type est lié à la différence de catégorie grammaticale et d’origine ((13a) et (13b)). Le deuxième type d’ambiguïté est lié à la forme écrite chinoise originale : (13b) et (13c) ont la même catégorie grammaticale et la même origine. Si cette forme *gajang* est écrite en caractères chinois dans un texte, cette ambiguïté peut être levée. Sinon, cette ambiguïté peut être enlevée par l’analyse syntaxique en utilisant les étiquettes attachées à chaque entrée : une étiquette *Hum* pour (13b) et une étiquette syntaxique *PRED* pour (13c) peuvent servir à

⁹ cf. *Kolyeo daehakgyo daehakgukyeo pyeonchansil* (Editorial Department of Korean for college students in Korea University) (1994).

les identifier¹⁰.

5. Les informations sémantiques

5-1. Noms humains, noms d'animaux et noms de plantes

Au niveau de la sémantique ou de la référence, un nom humain désigne un humain, un nom d'animal un animal et un nom de plante une plante. La distinction entre les noms humains, d'animaux, et de plantes est très importante pour la description syntaxique du lexique-grammaire et pour la combinaison avec certaines postpositions nominales comme nous l'avons mentionné dans la section 2 (pp. 69-72).

Il semble que la question en *nugu* « qui » et *mueot* « quoi » permette plus ou moins de distinguer les noms d'animaux des noms humains. Si le nom en position sujet ou complément d'objet direct répond au pronom interrogatif *nugu* « qui » ?, c'est un nom humain. Si le nom dans la même position répond à *mueot* « quoi » ?, c'est un nom non humain.

- (14) a. R : *eoje*, *Minu-ga* *hakgyo-eseo* *chingu-leul* *bo-at-da*
hier, Minu-**Nmtf** école-**Postp.eseo** ami-**Acc** voir-**Mpas-St**
(Hier, Minu a vu son ami à l'école)
Q : *Minu-ga* (*nugu* + **mueot*)-*leul* *bo-at-dago?*
Minu-**Nmtf** (qui + **quoi*)-**Acc** voir-**Mpas-Sti** ?
(Qu'est-ce que Minu a vu ?)
- b. R : *Minu-neun* *geoli-eseo* (*gae* + *namu* + *jip*)-*eul* *bo-at-da*
Minu-**Nmtf** rue-**Postp.eseo** (chien + arbre + maison)-**Acc** voir-**Mpas-St**
(Minu a vu (un chien + un arbre + une maison) dans la rue)
Q : *Minu-ga* *geoli-eseo* (**nugu* + *mueot*)-*leul* *bo-at-dago?*
Minu-**Nmtf** rue-**Postp.eseo** (**qui* + *quoi*)-**Acc** voir-**Mpas-Sti** ?
(Qu'est-ce que Minu a vu dans la rue ?)

Cependant, nous trouvons des cas auxquels le critère de la question en *nugu* (qui) et *mueot* (quoi) ne s'applique pas bien. *salam* « être humain » est incontestablement un nom humain. Mais dans le contexte suivant où il s'agit de l'emploi métonymique du

¹⁰ cf. Pour les détails sur les noms humains et les noms prédicatifs, voir les sections 5 et 7.

mot *salam* « être humain », le complément d'objet direct *salam* répond au pronom interrogatif *mueot* « quoi » ? :

- (15) R : *sikinjong-eun salam-eul meok-neunda*
 anthropophage-**Aux** être humain-**Acc** manger-**St**
 (Les anthropophages mangent des êtres humains)
 Q : *sikinjong-eun (*nugu + mueot)-leul meok-neundago ?*
 anthropophage-**Aux** (*qui + quoi)-**Acc** manger-**Sti** ?
 (Qu'est-ce que les anthropophages mangent ?)

Dans (16), nous remarquons que *gwisin* « démon » et *sin* « dieu » peuvent répondre, nous semble-t-il, à la fois aux pronoms interrogatifs *mueot* « quoi » ? et *nugu* « qui » ? :

- (16) R : *Minu-neun kkum-e (gwisin + sin)-eul bo-at-da*
 Minu-**Nmtf** rêve-**Postp.e** (démon + dieu)-**Acc** voir-**Mpas-St**
 (Minu a vu (un démon + le dieu) dans son rêve)
 Q : *Minu-ga kkum-e (nugu + mueot)-leul bo-at-dago?*
 Minu-**Nmtf** rêve-**Postp.e** (qui + quoi)-**Acc** voir-**Mpas-Sti** ?
 (Qu'est-ce que Minu a vu dans son rêve ?)

Il nous paraît que le critère « question en *nugu* (qui) et *mueot* (quoi) » n'est pas suffisant pour distinguer les noms humains des noms non humains. Ici, nous présentons deux critères formels qui permettent de distinguer de la façon rigoureuse les noms humains, d'animaux des noms inanimés et de plantes. Et nous classifions sémantiquement les séquences nominales figées à l'aide des deux critères formels suivants : celui de la compatibilité avec la *Postp.e*. et celui de la co-occurrence avec les noms classifieurs .

5-1-1. Compatibilité avec la *Postp.e*.

Les postpositions *-ege* (souvent langue écrite), *-hante* (souvent langue parlée) et *-e* sont les trois variantes de la forme canonique <-e>. La compatibilité des noms avec la postposition *-ege* ou la postposition *-e* nous permet de distinguer de façon rigoureuse les noms humains et d'animaux des noms inanimés et de plantes¹¹.

¹¹ cf. Seo J.S (1996), Lee I.S et Nam G. S (1987), Nam G.S et Go Y.G (1998), Heo Ung (1995), etc.

- (17) a. *Minu-ga chingu-(ege + hante + *e) chaek-eul ju-eot-da*
 Minu-**Nmtf** ami-(**Postp.e** + **Postp.e** + ***Postp.e**) livre-**Acc** donner-**Mpas-St**
 (Minu a donné un livre à son ami)
- b. *Minu-ga wonsungi-(ege + hante + *e) banana-leul ju-eot-da*
 Minu-**Nmtf** singe-(**Postp.e** + **Postp.e** + ***Postp.e**) banane-**Acc** donner-**Mpas-St**
 (Minu a donné une banane au singe)
- c. *Minu-neun (gwusin + sin)-(ege + hante + *e) gido-leul ha-nda*
 Minu-**Aux** (démon + dieu)-(**Postp.e** + **Postp.e** + ***Postp.e**) prière-**Acc** faire-**St**
 (Minu fait des prières au (démon + dieu))
- (18) a. *Minu-ga i namu-(*ege + *hante + e) mul-eul ju-eot-da*¹²
 Minu-**Nmtf** ce arbre-(**Postp.e** + **Postp.e** + ***Postp.e**) eau-**Acc** donner-**Mpas-St**
 (Minu a donné de l'eau à cet arbre)
- b. *Minu-ga i namu-(*ege + *hante + e) nakseo-leul ha-yeot-da*
 Minu-**Nmtf** ce arbre-(**Postp.e** + **Postp.e** + ***Postp.e**) griffonnage-**Acc** faire-**Mpas-St**
 (Minu a fait un griffonnage sur cet arbre)
- c. *Minu-ga i takja-(*ege + *hante + e) nakseo-leul ha-yeot-da*
 Minu-**Nmtf** ce table-(***Postp.e** + ***Postp.e** + **Postp.e**) griffonnage-**Acc** faire-**Mpas-St**
 (Minu a fait un griffonnage sur cette table)

Si le nom est compatible avec *-ege* ou *-hante*, c'est un nom humain ou d'animal. Si le nom n'est compatible qu'avec *-e*, c'est un nom inanimé ou de plante.

Pour les petits insectes comme les moustiques, les bactéries, les éphémères, etc. il nous semble que nous pouvons utiliser les trois formes *-ege*, *-hante* et *-e*, mais nous allons les distinguer selon notre deuxième critère dans la section 5-1-2.

5-1-2. Co-occurrence avec les noms classifieurs

Comme deuxième critère, nous proposons d'examiner la co-occurrence avec les noms classifieurs¹³. Par exemple, les noms humains peuvent avoir *myeong*, *in* ou *bun* (unité de personne) comme classifieurs. Pour les noms d'animaux, ils sont compatibles avec *su* ou *mali* (unité d'animal) et les noms de plantes avec *songi* ou *geulu* (pied).

¹² Dans un contexte spécifique de personnification des noms inanimés ou de plantes, on peut utiliser *-ege* ou *-hante*.

¹³ Nam J.S (1994) a utilisé ce critère dans le dictionnaire des noms simples.

- (19) a. *chingu daseot myeong* (ami cinq Nclas[personne] = cinq amis)
 b. *tokki se mali* (lapin trois Nclas[animal] = trois lapins)
 c. *podonamu du geulu* (vigne deux Nclas [plante] = deux pieds de vigne)

Ce critère est aussi utile pour distinguer les petits insectes :

- (20) a. (*mogi+ halusali*) *se mali* (moustique trois Nclas[animal] = trois (moustiques + éphémères))
 b. * *segyun se mali* (*bactérie trois Nclas[animal])

5-2. Les humains collectifs et les humains collectifs institutionnels

En coréen, il nous faut aussi distinguer deux sortes d'humains collectifs en fonction de la combinaison avec les postpositions casuelles : les noms humains collectifs (*peuple*) et les noms humains collectifs institutionnels (*gouvernement*). La postposition nominale <-eseo> ne marque le nominatif qu'avec les humains collectifs institutionnels¹⁴. Dans notre dictionnaire, nous avons codé ceux-ci *InstHum* ceux-là *CollHum*.

- (21) a. *wonsi minjok-(i + *eseo)* *geu il-e* *gaeipha-yeot-da*
 peuple primitif-(Nmtf +*Nmtf) ce affaire-**Postp.e** intervenir-**Mpas-St**
 (Un peuple primitif est intervenu dans ces affaires)
 b. *gongsu budae-(ga + eseo)* *geu il-e* *gaeipha-yeot-da*
 troupe parachutée-(Nmtf+Nmtf) ce affaire-**Postp.e** intervenir-**Mpas-St**
 (La troupe parachutée est intervenue dans ces affaires)

Quant aux postpositions <-e>, les noms humains collectifs se combinent avec les formes -ege et -hante et les noms humains institutionnels avec la forme -e comme les noms inanimés.

¹⁴ Dans le dictionnaire des postpositions nominales, nous avons traité les deux emplois différents de -eseo comme homonymes : Postp.eseo et le nominatif (cf. voir la section 3-3-1 du chapitre 4).

<i>na-neun</i>	<i>oneul</i>	<i>dosegwan-eseo</i>	<i>gongbu-leul</i>	<i>ha-yeot-da</i>
Je- Aux	aujourd'hui	bibliothèque- Postp.eseo	étude- Acc	faire- Mpas-St
(Aujourd'hui, j'ai étudié dans la bibliothèque)				

- (22) a. *daetonglyeong-i wonsi minjok-(ege + hante + *e) hanguk munhwajae-leul ju-eot-da*
 président-**Nmtf** peuple primitif-(**Postp.e** + **Postp.e** + ***Postp.e**) Corée (biens culturels)-
Acc donner-**Mpas-St**
 (Le président a donné les biens culturels de la Corée à un peuple primitif)
- b. *daetonglyeong-i gongsu budae-(*ege + *hante + e) keun seonmul-eul hasaha-yeot-da*
 président-**Nmtf** troupe parachutée-(***Postp.e** + ***Postp.e** + **Postp.e**) grand cadeau-**Acc**
 faire don de-**Mpas-St**
 (Le président a fait don d'un grand cadeau à la troupe parachutée)

Nous pouvons résumer les combinaisons avec les postpositions casuelles comme suit :

	Nom humain individuel Nom d'Animal	Nom inanimé Nom de plante	CollHum	InstHum
Postp.e	-hante / -ege	-e	-hante / -ege	-e
Nominatif	-ga /-i	-ga /-i	-ga /-i	-ga /-i, -eseo

Figure 1. Combinatoire avec la Postp.e et le nominatif

6. La fonction de sous-classifieur de N_1 et X dans N_1N_2 et XN

Le critère de l'acceptabilité de N_1N_2 est un (N + type de N_2 et XN est un (N + type de N) dans le chapitre 2 est utilisé pour l'identification des séquences nominales figées. Nous codifions cette information comme suit :

- (i) si N_1N_2 être un N_1 et N_1N_2 être un N_2 , le code est +12
- (ii) si N_1N_2 être un N_1 mais $*N_1N_2$ être un N_2 , le code est +1
- (iii) si $*N_1N_2$ être un N_1 mais N_1N_2 être un N_2 , le code est +2
- (iv) si $*N_1N_2$ être un N_1 et $*N_1N_2$ être un N_2 , le code est +0
- (v) si XN être un N , le code est +2
- (vi) si $*XN$ être un N , le code est +0
- (vii) X_1X_2 est codé +0, car $*X_1X_2$ être un X_2 ,
- (viii) si NX être un N , le code est +1

Voici un exemple de chaque cas :

- (23) a. *daehap jogae*, $N+NN+12$ « palourde » « coquillage » = « palourde »
 b. *iut sachon*, $N+NN+1$ « voisin » « cousin » = « voisin plus familier que des cousins »
 c. *chuli soseol*, $N+NN+2$ « déduction » « roman » = « roman policier »
 d. *jigak daejang*, $N+NN+0$ « retard » « chef » = « personne qui est toujours en retard »
 e. *minju jeongchi*, $N+XN+2$ « démocratique » « politique » = « politique démocratique »
 f. *gwagyeok bunja*, $N+XN+0$ « radical » « élément » = « extrémiste ou radical »
 g. *beseuteu seleo*, $N+XX+0$ « best » « seller » = « succès de librairie »
 h. *bitamin di*, $N+NX+1$ « vitamine » « D » = « vitamine D »

Etant donné que l'information sur la fonction de sous-classificateur permet de déduire quel nom joue sémantiquement le rôle de tête des séquences nominales (par exemple 1, 2, etc.), elle est évidemment utile pour la recherche d'information, l'extraction de mots-clés, etc.

7. Informations morpho-syntaxiques : les noms prédicatifs¹⁵

7-1. Définitions

Dans le cadre du lexique-grammaire¹⁶, l'étude des noms prédicatifs est motivée

¹⁵ Dans notre travail, nous excluons les noms humains prédicatifs en attendant des études sur les critères syntaxiques sur eux en coréen. Par exemple, en français les noms relationnels, que ce soit dans le domaine des relations de parenté (père, mère, cousin, etc.), dans le domaine spatial (voisin) ou professionnel (confrère, condisciple) sont des noms prédicatifs. Ils ont une syntaxe spécifique. Par exemple (G. Gross, 1992) :

**Je suis (un + E) frère*
Je suis le frère de Paul

En coréen, même si la notion relationnelle existe au niveau sémantique, il nous semble que les noms relationnels humains ne demandent pas distinctivement une syntaxe spécifique :

je-ga (geu ai-ui + E) *oppa-i-pnida*
 je-**Nmtf** (ce enfant-**Gén** + E) frère-être-**St**
 (Je suis le frère (de cet enfant + E))

cf. Han S. H (2000) a remarqué que les noms de métiers comme *byeonhosa* « avocat » peuvent être considérés comme les noms humains prédicatifs. Pour les détails, voir p. 8.

par le grand nombre de substantifs qui apparaissent comme prédicatifs. Dans la phrase suivante, *chanseong* « approbation » apparaît comme prédicatif, et *chanseong-eul hada* « donner son approbation » est sémantiquement équivalent au verbe à deux arguments *chanseonghada* « approuver » :

$$\begin{array}{ccccc} N_0 & W & N_{pred-Acc} & & hada \\ = N_0 & W & N_{pred-hada} & & \end{array}$$

- (24) a. *Minu-ga Ina-ui uigyeon-e daehayeo chanseong-eul ha-nda*
 Minu-**Nmtf** Ina-**Gén** opinion-**Postp.e daeha**¹⁷ approbation-**Acc** faire-**St**
 (Minu donne son approbation sur l'opinion de Ina)
- b. *Minu-ga Ina-ui uigyeon-e daehayeo chanseong-ha-nda*
 Minu-**Nmtf** Ina-**Gén** opinion-**Postp.e daeha** approuver-**St**
 (Minu approuve l'opinion de Ina)

Dans (24a), le nom *chanseong* joue le rôle du prédicat sémantique. C'est-à-dire que les arguments sémantiques sont liés au substantif *chanseong*, et non au verbe *hada*. Ce type de verbe est appelé verbe support : il est sémantiquement vide et sert de support grammatical. C'est un simple porteur de la morphologie de conjugaison verbale (par exemple, du mode, du temps, de l'aspect, etc.). Ce fait est lié à la possibilité de former un groupe nominal (24d) à partir d'une phrase simple à verbe support (24a), ici par réduction de relative à partir de (24c), selon la règle [Réd Vsup]. C'est une caractéristique syntaxique des noms prédicatifs.

[Relativation]

- c. *[Minu-ga Ina-ui uigyeon-e daehayeo ha-n chanseong-i] Ina-leul gippeuge ha-yeot-da*
 [Minu-**Nmtf** Ina-**Gén** opinion-**Postp.e daeha** faire-**Sfd** approbation-**Nmtf**] Ina-**Acc**
 faire plaisir-**Mpas-St**
 ([L'approbation que Minu a donnée sur l'opinion de Ina] lui a fait plaisir)

¹⁶ Dans les dictionnaires du type DELA du LADL, les noms abstraits sont considérés comme des noms prédicatifs, et ils ont fait l'objet d'études syntaxiques au LADL.

¹⁷ Selon Han S.H (2000, pp. 74-75), la postposition *-e daehayeo* est une postposition composée. Elle est constituée d'une postposition simple *-e* et d'un verbe *dae-hada* suffixé en *-hada*. Ce verbe est accompagné du suffixe conjonctif (*Sconj*) *-yeo* : *-e daeha-yeo* « *Postp.e* » « *V-Sconj* » = « à propos de ». Le remplacement du *Sconj* = : *yeo* par le suffixe déterminatif *-n* (*Sfd*) conduit à une variante dans le complément nominal : *-e daeha-n* « *Postp.e* » « *V-Sfd* » = « à propos de ».

[Réd Vsup]

- d. [Minu-ui Ina-ui uigyeon-e daeha-n chanceong-i] Ina-leul gippeuge ha-yeot-da
[Minu-**Gén** Ina-**Gén** opinion-**Postp.e daeha** approbation-**Nmtf**] Ina-**Acc** faire plaisir-
Mpas-St
([L'approbation de Minu sur l'opinion de Ina] lui a fait plaisir)

7-2. Les noms compacts prédicatifs (*Npred*)

Au point de vue morpho-syntaxique, nous observons deux types de noms prédicatifs : ceux qui ont une construction verbale associée (*Npred* = : *V-n*) et ceux qui n'en ont pas (*Npred* n'est pas un *V-n*)¹⁸.

7-2-1. *Npred* (= : *V-n*)¹⁹

D'après Z. Harris, les nominalisations²⁰ sont des relations transformationnelles non orientées et elles ne correspondent pas à une règle transformant une phrase en un groupe nominal. La nominalisation se définit comme suit (A. Monceaux, 1993) :

L'approche adoptée au LADL est inspirée des travaux de Z.S. Harris, c'est-à-dire que la nominalisation est envisagée en termes d'équivalence entre deux phrases, et non plus entre une phrase et un groupe nominal. La nominalisation met en œuvre une des classes d'opérateurs pour dériver une phrase comportant un nom dérivé d'une autre comportant un verbe ou un adjectif. Elle peut alors être conçue comme une simple relation d'équivalence ou de synonymie entre deux phrases, et non pas comme une transformation dans le sens génératif du terme. La relation entre les phrases n'est pas orientée, elle est notée par signe d'égalité :

Luc décide d'en finir
=
Luc prend la décision d'en fuir

¹⁸ Nous notons *V-n* un substantif morpho-syntaxiquement associé à un verbe. Or, il est nécessaire de distinguer deux types de nominalisations en coréen : la nominalisation produisant un nom prédicatif (*V-n*) comportant parfois un suffixe nominal *-eum*, *-gi*, *-i*, etc., et la nominalisation systématique par les suffixes de nominalisation complémentaires *SfnComp* (:= *-eum* et *-gi*), qui introduit une complétive dans une phrase et est comparable à celle en *-ing* en anglais. Dans ce travail, nous ne traitons pas la relation de nominalisation par *SfnComp*. Mais nous indiquons ici que les noms prédicatifs (*V-n*) sont morphologiquement semblables aux formes nominalisées des complétives en *SfnComp*.

¹⁹ Nous notons *Npred* les noms compacts prédicatifs.

²⁰ cf. J. Giry (1978), G. Gross (1989), etc.

Max espère que tu viendras
 = *Max a l'espoir que tu viendras*

En coréen, la nominalisation qui met en relation des verbes avec des *Npred* (=V-n) peut se présenter sous deux formes : l'une met en jeu un suffixe nominal (-*eum*, -*gi*, -*i*, -*lam*, etc.) pour la plupart des *Npred* proprement coréens, et par l'autre suffixe verbal -*hada* pour les *Npred* sino-coréens et empruntés, et pour quelques types de *Npred* proprement coréens. Dans les deux cas, la construction nominale comporte un *Vsup*.

La nominalisation à suffixe nominal -*eum*, -*gi*, -*i*, -*lam*, etc. n'est possible que pour les verbes (ou les adjectifs) simples proprement coréens.

- (25) a. *Ina-ga* *Minu-leul* *dou-at-da*
 Ina-**Nmtf** Minu-**Acc** aider-**Mpas-St**
 (Ina a aidé Minu)
- = b. *Ina-ga* *Minu-ege* *doum-eul* *ju-eot-da*
 Ina-**Nmtf** Minu-**Postp.ege** aide-**Acc** donner-**Mpas-St**
 (Ina a donné de l'aide à Minu)
- (26) a. *Minu-wa* *Ina-ga* *ssau-eot-da*
 Minu- **Postp.wa** Ina-**Nmtf** se disputer-**Mpas-St**
 (Minu et Ina se sont disputés)
- b. *Minu-wa* *Ina-ga* *ssaum-eul* *ha-yeot-da*
 Minu- **Postp.wa** Ina-**Nmtf** dispute-**Acc** faire-**Mpas-St**
 (Minu et Ina ont eu une dispute)

Les noms *doum* (aide) et *ssaum* (dispute) dans (25b) et (26b) sont des noms déverbaux qui comportent une base verbale et le suffixe de nominalisation -*eum*. Les phrases (b) sont en relation de nominalisation avec les phrases (a) à verbe simple.

La majorité des noms prédicatifs sino-coréens sont associés morpho-syntaxiquement à l'adjectif ou au verbe correspondants *Npred-hada*, et ne comportent aucun des suffixes nominaux -*eum*, -*gi*, -*i*, etc. Au contraire, ce sont les verbes dénominaux qui comportent le suffixe verbal -*hada*.

- (27) a. *Ina-ga* *Minu-leul* *gonggyeok-eul* *ha-yeot-da*
 Ina-**Nmtf** Minu-**Acc** attaque-**Acc** faire-**Mpas-St**

(Ina a mené une attaque contre Minu)

b. *Ina-ga Minu-leul gonggyeokha-yeot-da*

Ina-**Nmtf** Minu-**Acc** attaquer-**Mpas-St**

(Ina a attaqué Minu)

Les deux phrases présentent donc un phénomène intéressant de parallélisme entre deux combinaisons de *Npred* et de *hada*:

$N_0 \quad W \quad Npred-Acc \quad hada = N_0 \quad W \quad Npred-hada$

Han S.H (2000) appelle *formation de verbes complexes N-hada* cette nominalisation.

Ce phénomène dérivationnel est aussi observé pour les noms empruntés à d'autres langues. Par exemple, *angketeu* « enquête » est un nom prédicatif emprunté au français. Le verbe *angketeu-hada* « enquêter » qui comporte le suffixe verbal *-hada* est lié au nom *angketeu*.

(28) a. *Ina-ga geoli-eseo angketeu-leul ha-yeot-da*

Ina-**Nmtf** rue-**Postp.eseo** enquête-**Acc** faire-**Mpas-St**

(Ina a fait une enquête dans la rue)

b. *Ina-ga geoli-eseo angketeuha-yeot-da*

Ina-**Nmtf** rue-**Postp.eseo** enquêter-**Mpas-St**

(Ina a enquêté sur l'opinion publique dans la rue)

Enfin, quant aux noms prédicatifs proprement coréens, nous observons deux types de *Npred* associés à *Npred-hada*. Par exemple, le *V-n* à suffixe nominal de (26b) peut se combiner avec le suffixe verbal *-hada* pour former un nouveau verbe *V-n-hada*. Parallèlement aux paires (26a) et (26b), on a ainsi la phrase équivalente suivante :

(26) c. *Minu-wa Ina-ga ssaumha-yeot-da*

Minu- **Postp.wa** Ina-**Nmtf** se disputer-**Mpas-St**

(Minu et Ina se sont disputés)

Les verbes qui présentent ces trois formes (verbe sans suffixe, nom à suffixe, verbe à double suffixe) ne sont pas nombreux. Les *Npred* déverbaux (*V-n*) sont reliés, d'une part à un verbe simple par la relation de nominalisation et permettent, d'autre part, la formation d'un verbe complexe *V-n-hada*.

L'autre type d'exemple fait intervenir le suffixe *-jil*²¹ qui donne une valeur prédicative aux noms avec lesquels il se combine. *geolle* « serpillière » est un nom concret d'outil et *-jil* désigne l'utilisation de cet outil. Lorsque *-jil* se combine avec des noms d'outils *Nconc*, *Nconc-jil* désigne l'utilisation de *Nconc*. Ainsi, *geolle-jil* « action de passer la serpillière » est un nom prédicatif en *-jil*.

- (29) a. *Ina-ga tail-e geollejil-eul ha-yeot-da*
 Ina-**Nmtf** carrelage-**Postp.e** (action de passer la serpillière)-**Acc** faire-**Mpas-St**
 (Ina a passé la serpillière sur le carrelage)
- b. *Ina-ga tail-e geollejilha-yeot-da*
 Ina-**Nmtf** carrelage-**Postp.e** (passer la serpillière)-**Mpas-St**
 (Ina a passé la serpillière sur le carrelage)

Le suffixe *-jil* est également compatible avec certains *Npred* intrinsèquement prédicatifs et exprimant une action :

- (30) a. *Minu-wa Jinsu-ga ssaumjil-eul ha-yeot-da*
 Minu- **Postp.wa** Jinsu-**Nmtf** (action de se disputer)-**Acc** faire-**Mpas-St**
 (Minu et Jinsu ont eu une dispute)
- b. *Minu-wa Jinsu-ga ssaumjilha-yeot-da*
 Minu-**Postp.wa** Jinsu-**Nmtf** (action de se disputer)-**Mpas-St**
 (Minu et Jinsu se sont disputés)

Le nombre de noms prédicatifs *Npred* proprement coréens associés à *Npred-hada* est limité. Nous allons finir cette section en présentant quelques cas discutables de radicaux proprement coréens (*Rx*). Certains radicaux *Rx* forment des adjectifs de la forme *Rx-hada* avec le suffixe *-hada* dans (31a) et (32a), et ils sont parfois considérés comme des noms prédicatifs dans (31b) et (32b), où ils se combinent avec un verbe support autre que *-hada*.²²

²¹ Han S.H (1996) a appelé *-jil* un suffixe prédicatif. Elle a remarqué que les noms dérivés en *-jil* ont les points communs suivants :

- Ce sont des noms abstraits qui expriment une action ou une activité.
- Ils se combinent avec le verbe *hada* (faire), en jouant le rôle du prédicat dans des phrases dont le sujet est humain.

²² cf. Hong CH.S (1999) a considéré les radicaux non autonomes *bujileon* (diligence) et *jeomjan* (dignité) comme des noms prédicatifs qui ont une distribution assez restreinte.

- (31) a. *Ina-ga bujileon-ha-da*
 Ina-**Nmtf** diligent-être-**St**
 (Ina est diligente)
- b. *Ina-ga bujileon-eul (tteon + *ha)-da*
 Ina-**Nmtf** diligence-**Acc** (montrer + *faire)-**St**
 (Ina (montre + *fait) de la diligence)
- (32) a. *Minu-ga jeomjan-ha-da*
 Minu-**Nmtf** l'air digne-avoir-**St**
 (Minu a l'air digne)
- b. *Minu-ga jeomjan-eul (ppaeda + *hada)*
 Minu-**Nmtf** dignité -**Acc** (montrer + *faire)
 (Minu (montre + *fait) de la dignité)

Or, il n'est pas possible de former des groupes nominaux à partir des phrases simples (31b) et (32b) :

- (31) c. *[*Minu-ui bujileon-i*] *Ina-leul gippeuge ha-yeot-da*
 [Minu-**Gén** diligent-**Nmtf**] Ina-**Acc** faire plaisir-**Mpas-St**
 c'. [*Minu-ui bujileon-ham-i*] *Ina-leul gippeuge ha-yeot-da*
 [Minu-**Gén** diligence-**Nmtf**] Ina-**Acc** faire plaisir-**Mpas-St**
 (La diligence de Minu a fait plaisir à Ina)
- (32) c. *[*Minu-ui jeomjan-i*] *Ina-leul gippeuge ha-yeot-da*
 [Minu-**Gén** digne-**Nmtf**] Ina-**Acc** faire plaisir-**Mpas-St**
 c'. [*Minu-ui jeomjan-ham-i*] *Ina-leul gippeuge ha-yeot-da*
 [Minu-**Gén** dignité-**Nmtf**] Ina-**Acc** faire plaisir-**Mpas-St**
 (La diligence de Minu a fait plaisir à Ina)

La règle [Réd Vsup] n'est pas applicable aux (31b) et (32b). *debujileon* « diligence » et *jeomjan* « dignité » ne sont pas des noms prédicatifs et ils sont considérés comme des expressions figées : *bujileon-eul tteolda* « montrer de la diligence » et *jeomjan-eul ppaeda* « montrer de la dignité ».

7-2-2. Les *Npred* qui ne sont pas des *V-n*

Ici, nous nous contentons de citer quelques exemples de noms prédicatifs non

associés à des verbes. Les exemples de (33) et (34) sont des *Npred* auxquels on ne peut pas faire correspondre un verbe par suffixation en *-hada*. Ces *Npred* s'emploient avec des verbes supports comme *batda* « subir, recevoir » ou *jeojileuda* « commettre » qui ne servent pas de suffixes.

- (33) a. *Ina-ga chingu-deul-lo-buteo (nolim + moyok)-eul bat-at-da*
 Ina-**Nmtf** chingu-**Mplu-Postp.lo-Postp.buteo** (moquerie +injure)-**Acc** subir-**Mpas-St**
 (Ina a subi une (moquerie + injure) des amis)
- b. **Ina-ga chingu-deul-lo-buteo (nolim + moyok)-bat-at-da*
 *Ina-**Nmtf** ami-**Mplu-Postp.lo-Postp.buteo** (moquerie +injure)-subir-**Mpas-St**
- c. **Ina-ga chingu-deul-lo-buteo (nolim + moyok)-eul ha-yeot-da*
 *Ina-**Nmtf** ami-**Mplu-Postp.lo-Postp.buteo** (moquerie +injure)-**Acc** faire-**Mpas-St**
- d. **Ina-ga chingu-deul-lo-buteo (nolim + moyok)-ha-yeot-da*
 *Ina-**Nmtf** ami-**Mplu-Postp.lo-Postp.bute** (moquerie+injure)-faire-**Mpas-St**
- (34) a. *Minu-ga salinjoe-leul jeojill-eot-da*
 Minu-**Nmtf** meurtre-**Acc** commettre-**Mpas-St**
 (Minu a commis un meurtre)
- b. **Minu-ga salinjoe-jeojill-eot-da*
 *Minu-**Nmtf** meurtre-commettre-**Mpas-St**
- c. **Minu-ga salinjoe-leul ha-yeot-da*
 *Minu-**Nmtf** meurtre-**Acc** faire-**Mas-St**
- d. **Minu-ga salinjoe-ha-yeot-da*
 *Minu-**Nmtf** meurtre-faire-**Mas-St**

Nous récapitulons ici notre classification des noms compacts prédicatifs (*Npred*) :

Orig.	<i>Npred</i> =: <i>V-n</i>	<i>Npred</i> n'est pas un <i>V-n</i>
pk	(26) N_0 W <i>ssauda</i> = N_0 W <i>ssaum</i> -(Acc+E) <i>hada</i> (N_0 se disputer W = N_0 avoir une dispute W) (29) N_0 W <i>geollejihada</i> = N_0 W <i>geollejil</i> -Acc <i>hada</i> (N_0 passer la serpillière W = N_0 fait l'action de passer la serpillière W)	(33) * N_0 W <i>nolimhada</i> N_0 W <i>nolim</i> -Acc <i>batda</i> (N_0 suivre la moquerie W)
sk	(27) N_0 N_I W <i>gonggyeokhada</i> = N_0 N_I W <i>gonggyeok</i> -Acc <i>hada</i> (N_0 attaquer N_I W = N_0 mener une attaque contre N_I W)	(34) * N_0 W <i>salinjoehada</i> N_0 W <i>salinjoe</i> -Acc <i>jeojileuda</i> (N_0 commettre un meurtre W)

Figure 2. Les noms compacts prédicatifs

7-3. Les noms figés prédicatifs (*NFpred*)²³

Dans le cadre de la phrase simple, les noms figés se comportent, par rapport aux verbes supports, de la même façon que les noms compacts. Si nous remarquons une propriété distinctive des noms figés prédicatifs (*NFpred*) par rapport aux *Npred*, c'est que la prédicativité de la majorité des *NFpred* est, en général, liée à la prédicativité de leur nom tête.

Toutefois, les *NFpred* se subdivisent en trois catégories : les *NFpred* associés à une construction verbale figée par nominalisation, les *NFpred* dont seul le nom tête est morpho-syntaxiquement associé à un verbe, et les *NFpred* qui n'ont aucune construction verbale associée.

7-3-1. *NFpred* (= *V-n*)

Les *NFpred* associés à des constructions verbales sont des cas de nominalisation de phrases figées, soit proprement coréennes, soit sino-coréennes.

- (35) a. *Miun-neun* *bang-eseo* *mulgunamu-leul* *seo-nda*
Minu-**Aux** chambre-**Postp.eseo** (faire le poirier)-**St**
(Minu fait le poirier dans la chambre)
- = b. *Muniu-neun* *bang-eseo* *mulgunamu seo-gi-leul* *ha-nda*
Minu-**Aux** chambre-**Postp. eseo** (faire le poirier-**Sfn** = **V-n**)-**Acc** faire-**St**
(Minu fait le poirier dans la chambre)

²³ Nous notons *NFpred* les noms figés prédicatifs.

L'exemple (35a) est une phrase figée *mulgunamu*-(*Acc + E*) *seoda* et (35b) est une nominalisation de phrase figée avec suffixe nominal *-gi*. De plus, comme (26c) (p. 86), le *V-n* de (35) peut se combiner avec le suffixe *-hada*, ce qui produit la phrase équivalente suivante :

- (35) c. *Muniu-neun bang-eseo mulgunamu seo-gi-ha-nda*
 Minu-**Aux** chambre-**Postp.e** *eseo* (faire le poirier-**Sfn = V-n**)-faire-**St**
 (Minu fait le poirier dans la chambre)

Le nombre d'exemples de ce type parmi les *NFpred* est assez limité. Plusieurs phrases figées sino-coréennes à quatre syllabes appartiennent à ce type de *NFpred* (=V-n) :

- (36) a. *Ina-ga sinbunim-kke gobaek seongsa-leul ha-yeot-da*
 Ina-**Nmtf** prêtre-**Postp.e** (confession religieuse)-**Acc** faire-**Mpas-St**
 (Ina a fait une confession au prêtre ; au sens religieux)
- b. *Ina-ga sinbunim-kke gobaek seongsa-ha-yeot-da*
 Ina-**Nmtf** prêtre-**Postp.e** (se confesser)-**Mpas-St**
 (Ina s'est confessée au prêtre)
- c. **Ina-ga sinbunim-kke seongsa-ha-yeot-da*
 *Ina-**Nmtf** prêtre-**Postp.e** (rite religieux)-faire-**Mpas-St**
- d. *Ina-ga sinbunim-kke gobaek-eul ha-yeot-da*
 Ina-**Nmtf** prêtre-**Postp.e** aveu-**Acc** faire-**Mpas-St**
 (Ina a fait un aveu au prêtre)

Le sens de *gobaek seongsa* « confession - rite religieux = confession religieuse (au sens religieux) » n'est pas calculable à partir du sens de ses éléments. De plus, (36d) est une phrase acceptable, mais le sens est différent : il ne s'agit plus du rite catholique.

7-3-2. *NFpred* à nom tête déverbal

Maintenant, nous allons considérer les *NFpred* dont le nom tête est morpho-syntaxiquement associé à une construction verbale. La plupart de ces *NFpred* sont sino-coréens et ils sont très nombreux. Commençons par des exemples sino-coréens des structures *XN* et *N₁N₂*.

- (37) a. *Minu-ga gukje gyeolhon-eul ha-nda*
 Minu-**Nmtf** mariage international-**Acc** faire-**St**
 (Minu se marie avec un étranger)
- a'. *Minu-ga gyeolhon-eul ha-nda*
 Minu-**Nmtf** mariage-**Acc** faire-**St**
 (Minu se marie)
- = a''. *Minu-ga gyeolhon-ha-nda*
 Minu-**Nmtf** se marier-**St**
 (Minu se marie)
- b. *jikwon-deul-i dansik tujaeng-eul ha-nda*
 employeur-**Mplu-Nmtf** (grève de la faim)-**Acc** faire-**St**
 (Les employés font la grève de la faim)
- b'. *jikwon-deul-i tujaeng-eul ha-nda*
 employeur-**Mplu-Nmtf** grève-**Acc** faire-**St**
 (Les employés font une grève)
- = b''. *jikwon-deul-i tujaeng-ha-nda*
 employeur-**Mplu-Nmtf** faire une grève-**St**
 (Les employés font une grève)

Dans (37a), *gukje* (X) « international » qui est un élément sans catégorie grammaticale fonctionne comme un sous-classificateur de *N*, et *gyeolhon* « mariage » est un *Npred* (= :V-n) à verbe support *-hada*. Dans (37b), *dansik* (*N₁*) joue le rôle de sous-classificateur de *N₂* (*Npred*). *X Npred* et *N Npred* sont abrégés en *Npred*. Comme nous le voyons dans les exemples de (37a') - (37a'') et (37b') - (37b''), la prédicativité de noms figés dans (37a) et (37b) est liée à celle de leurs noms têtes. Ce fait est aussi observé dans l'exemple (38) de *NFpred* proprement coréen à nom tête déverbal admettant *hada* comme *Vsup* :

- (38) a. *Ina-ga tail-e gileum geollejil-eul ha-yeot-da*
 Ina-**Nmtf** carrelage-**Postp.e** (action de passer le chiffon gras)-**Acc** faire-**Mpas-St**
 (Ina a passé le chiffon gras sur le carrelage)
- b. *Ina-ga tail-e geollejil-eul ha-nda*
 Ina-**Nmtf** carrelage-**Postp.e** (action de passer la serpillière)-**Acc** faire-**St**
 (Ina a passé la serpillière sur le carrelage)
- = b'. *Ina-ga tail-e geollejilhan-da*

Ina-**Nmtf** carrelage-**Postp.e** (passer la serpillière) -**St**
 (Ina passe la serpillière sur le carrelage)

Considérons un autre exemple de *NFpred* proprement coréen à nom tête déverbal par nominalisation en *-i* :

- (39) a. **Minu-ga* *kkokdukgaksi-(leul + E)* *nol-da*
 Minu-Nmtf* marionnette-(Acc + E**) jouer-**St**
 b. *Minu-ga* *kkokdukgaksi* *nol-i-leul* *ha-nda*
 Minu-Nmtf marionnette jouer-**Sfn-Acc** faire-**St**
 (Minu faire un spectacle de marionnettes)
 b'. *Minu-ga* *nol-i-leul* *ha-nda*
 Minu-Nmtf jouer-**Sfn-Acc** faire-**St**
 (Minu faire un jeu)
 = b''. *Minu-ga* *no-nda*
 Minu-Nmtf jouer-**St**
 (Minu joue)

Il n'existe aucune phrase équivalente à *kkokdukgaksi noli* « jeu de marionnettes » de (39b) qui soit du type **kkokdukgaksi-(Acc+E) nolda* « *jouer des marionnettes ». En fait, l'exemple (39b) est un *NFpred* proprement coréen dont la prédicativité est attachée au nom tête en l'absence du premier élément. (39b') est une nominalisation de (39b'') par un suffixe nominal *-i*.

La prédicativité des noms figés en coréen est souvent liée à celle de leurs noms têtes. Cette conclusion est comparable à celle de A. Monceaux (1993) : en français, les noms composés *NA* de type *Npred N-a*²⁴ où *N-a* fonctionne comme un sous-

²⁴ Dans *Npred N-a*, *N-a* signifie un adjectif de relation (ou un pseudo-adjectif). Il nous semble que la structure *X Npred* en coréen a certaines similarités avec celle de *Npred N-a* en français. *N-a* a les propriétés suivantes (A. Monceaux, 1993) :

- une relation morphologique avec un nom
- la non-prédicativité, c'est-à-dire l'interdiction de la position attribut, ou bien la prédicativité contrainte de d'adjectif. Par exemple : *une décision gouvernementale* = **Cette décision est gouvernementale*
- l'interdiction de la modification par un adverbe de degré, y compris dans les phrases attributives recevant une interprétation contrastive : **Cette allergie est très pollinique*

Cet adjectif de relation fonctionne comme un sous-classificateur de *Npred* dans *Npred N-a* où l'interprétation de « *Npred N-a est (un N + un type de N)* » est acceptable. L'adjectif de relation est paraphrasable par un complément de (*E + LE*) *N*. Par exemple : *une décision (gouvernementale + du*

classificateur de *Npred* sont abrégables en *Npred*. Il nous semble qu'en coréen, les noms figés à espacement facultatif de type *N Npred* ou *X Npred*, où *N* et *X* fonctionnent comme un sous-classificateur de *Npred*, sont abrégables en *Npred*.

7-3-3. Les *NFpred* qui ne sont pas des *V-n*

Ces *NFpred* ne sont morpho-syntaxiquement associés ni à des phrases figées ni à une construction verbale reliée à leur nom tête, mais ils peuvent former des constructions avec d'autres verbes supports comme *jeojileuda* « commettre ». Leur nom tête (e.i., *salinjoe* « meurtre »)²⁵ peut être un *Npred* mais n'a pas de construction verbale associée.

- (40) a. *Minu-ga gangdo salinjoe²⁶-leul jeojill-eot-da*
 Minu-**Nmtf** (délit de cambriolage et de meurtre)-**Acc** commettre-**Mas-St**
 (Minu a commis un délit de cambriolage et de meurtre)
- b.**Minu-ga gangdo salinjoe-jeojill-eot-da*
 *Minu-**Nmtf** (délit de cambriolage et de meurtre)-commettre-**Mas-St**
- c.**Minu-ga gangdo salinjoe-leul ha-yeot-da*
 *Minu-**Nmtf** (délit de cambriolage et de meurtre)-**Acc** faire-**Mas-St**
- d.**Minu-ga gangdo salinjoe-ha-yeot-da*
 *Minu-**Nmtf** (délit de cambriolage et de meurtre)-faire-**Mas-St**

Les *NFpred* proprement coréens qui ne sont pas des *V-n* sont très peu. Nous présentons un exemple dans lequel le *NFpred* proprement coréen, *hoeoli balam* (tourbillon - vent = tourbillon de vent) est le sujet de son verbe support.

- (41) a. *hoeoli balam-i chi-nda*
 tourbillon de vent-**Nmtf** faire-**St**
 (Il tourbillonne)
- b.* *hoeoli balam-chi-nda*
 * tourbillon de vent-faire-**St**

gouvernement) (*Npred N-a* + *N₁ de Le N₂*) où les exemples de *N₁ de Le N₂* sont des groupes nominaux libres et les exemples de *Npred N-a* sont figés en français. Les séquences nominales *N₁ de N₂* et *X-jeok* (ou *X-han*) *N* sont des groupes nominaux libres, tandis que les séquences nominales *Npred N-a* et *X Npred* sont des séquences nominales figées.

²⁵ cf. voir l'exemple (34), p. 89.

²⁶ C'est une appellation légale, mais très courante.

- c.* *hoeoli balam-i ha-nda*
 * tourbillon de vent-**Nmtf** faire-**St**
- d. * *hoeoli balam ha-nda*
 * tourbillon de vent-faire-**St**

Nous récapitulons ici notre classification des noms figés prédicatifs (*NFpred*) :

Orig.	<i>NFpred</i> =:V-n (nominalisation de phrase figée)	<i>NFpred</i> à nom tête deverbal	<i>NFpred</i> n'est pas un V-n
pk	(35) <i>N₀ W mulgunamu- (Acc+E) seoda</i> = <i>N₀ W mulgunamu seogi- (Acc+E) hada</i> (<i>N₀</i> faire le poirier W)	(38) <i>N₀ W gileum geollejil-Acc hada</i> (<i>N₀</i> W passer le chiffon gras) cf. <i>N₀ W geollejil-(Acc+E) hada</i> = <i>N₀</i> (<i>N₀</i> passer la serpillière W) (39) * <i>N₀ W kkokdukgaksi nolda</i> <i>N₀ W kkokdukgaksi noli-Acc hada</i> (<i>N₀</i> W faire un spectacle de marionnettes) cf. <i>N₀ W nolda</i> = <i>N₀ W noli-(Acc+E) hada</i> (<i>N₀</i> W jouer = <i>N₀</i> W faire un jeu)	(41) * <i>N₀</i> (= : <i>NFpred</i>) W <i>hada</i> <i>hoeoli balam-Nmtf W</i> <i>chida</i> (Il tourbillonne W)
sk	(36) <i>N₀ W gobaek seongsa-(Acc+E) hada</i> (<i>N₀</i> faire une confession W)	(37b) <i>N₀ W dansik tujaeng-Acc hada</i> (<i>N₀</i> W faire la grève de la faim) cf. <i>N₀ W tujaeng-Acc hada</i> = <i>N₀ W tujaenghada</i> (<i>N₀</i> W faire une grève)	(40) * <i>N₀ W gangdo salinjoe- (Acc+E) hada</i> <i>N₀ W gangdo salinjoe- Acc jeojilleuda</i> (<i>N₀</i> W commettre un délit de cambriolage et de meurtre)

Figure 3. Les noms figés prédicatifs

7-4. Quelques remarques sur la relation de paraphrase dans la structure *X Npred*²⁷

Revenons au problème de la relation de paraphrase entre noms figés de structure *X Npred* et groupes nominaux libres. Nous avons brièvement évoqué ce problème dans la section 2-4-2 du chapitre 2. Nous distinguerons deux types de *X Npred* : le type A et le type B, suivant que le suffixe insérable entre *X* et *Npred* est -*jeokin* ou -*han*.

7-4-1. Type A : *Rx(-jeokin + *-han + E) Npred*

²⁷ Pour l'instant, nous ne considérons que les noms prédicatifs à verbe support -*hada*.

Reprenons l'exemple (31) de la section 2-4-2 dans le chapitre 2.

- (42) a. *minjujeok-i-n jeongchi* (A-Sfd Npred)(libre)
 (démocratique- être-Sfd - politique = politique qui a un caractère démocratique)
 = b. *minju jeongchi* (XNpred)(figé)(démocratique - politique = politique
 démocratique)

Les exemples (42a) et (42b) sont approximativement synonymes. Nous avons vérifié les contraintes de compatibilité entre *Rx* et *-jeok(in)*, *Rx-jeok* et *Npred*, et *Rx-jeokin* et *Npred* par l'examen des exemples (27) - (30) du chapitre 2. Par conséquent, nous pouvons formuler l'hypothèse suivante :

- Si le *Npred* déterminés par le *X* est sino-coréen ou emprunté, le *X Npred* (figé) est paraphrasable à la fois par *Rx-jeokin Npred* et *Rx-jeok Npred* (libres).
- Si le *Npred* déterminé est proprement coréen, le *X Npred* (figé) n'est paraphrasable que par *Rx-jeokin Npred* (libre).

Le morphème proprement coréen *-i-n* joue un rôle d'intermédiaire entre l'élément sino-coréen (*X-jeok*) et l'élément proprement coréen (*N*), et la séquence *X-jeok N* est peu naturelle. Nous vérifions notre hypothèse dans les exemples (43).

- (43) a. *minju jeongchi* (sk **sk**)(X Npred)(figé)
 (démocratique - politique = politique démocratique)
 = *minju-(jeokin + jeok) jeongchi* ((A-Sfd + N) N)(libre)
 (être démocratique - politique = politique démocratique)
- b. *ganjeop gui* (sk **pk**)(X Npred)(figé)
 (indirect - cuisson = cuisson au feu ou à la braise dans laquelle les aliments ne sont pas directement exposés à la source de chaleur)
 = *ganjeop-(jeokin + *jeok) gui* ((A-Sfd + N) N)(libre)
 (être indirect - cuisson = cuisson indirecte)

Pour les variantes de la structure *X N*, nous devons considérer le statut prédicatif du *N* déterminé par le *X*. Si les *X N* du type A : *Rx(-jeokin + *-han + E)* sont figés et que *N* est un nom prédicatif, ils sont paraphrasables par *Rx-jeokin N* ou *Rx-jeok N*. Sinon, les *X*

N ne le sont pas (cf. voir le sens différent entre (32a) et (32b) dans le chapitre 2, pp. 51-52). Cette variation est extensible aux noms figés N_1N_2 où N_2 est un nom prédicatif :

- (44) a. *munhwa saenghwal* (*sk sk*)(*N Npred*)(figé)(culture - vie = vie culturelle)
 = b. *munhwa-(jeok + jeokin) saenghwal* (*N-(jeok + jeokin) Npred*)(libre)
 (être culturel- vie = vie culturelle)

L'exemple (44) *munhwa saenghwal* (N_1N_2) est compatible avec *-jeok(in)*, mais pas avec *-han*. Les noms N_1N_2 figés sont aussi paraphrasables par N_1 -*jeokin* N_2 et N_1 -*jeok* N_2 .

En revanche, la séquence nominale libre à structure *N Npred* (45a) ne permet pas une telle relation de paraphrase :

- (45) a. *gwahak yeongu* (*sk sk*)(*N Npred*)(libre)(science - étude = étude de la science)
 ≠ b. *gwahak-jeok yeongu* (*N-(jeok + jeokin) Npred*)(libre)
 (être scientifique-étude = étude scientifique)

Le sens des séquences nominales libres (45a) et (45b) est différent : (45a), dans lequel le premier nom est un complément d'objet direct du nom prédicatif (*yeongu*), signifie que l'objet de l'étude est la science, alors que (45b) qualifie la façon dont l'étude est menée.

Maintenant, nous allons examiner la dérivation adverbiale avec la postposition *-eulo* (*Rx-jeok-eulo*). Selon Nam J.S (1996), *-eulo* ne concerne que des adjectifs en *-ida*, qui n'acceptent aucune autre forme adverbiale que celle formée avec la postposition *-eulo*. Reprenons l'exemple (43a). Nous constatons une autre relation de paraphrase entre la construction adjectivale et la construction adverbiale.

- (46) a. *minju jeongchi* (*X Npred*)(figé)
 (démocratique - politique = politique démocratique)
 a'. *minju-(jeokin + jeok) jeongchi* ((*A-Sfd + N*) *Npred*)(libre)
 a''. *minju-jeok-eulo jeongchi* (*Rx-jeok-eulo Npred*)
 (démocratiquement - politique = (faire) de la politique démocratiquement)
 b. *Minu-neun minju-(E + jeokin + jeok) jeongchi-leul ha-nda*
 Minu-**Aux** démocratique politique-**Acc** faire-**St**
 (Minu fait de la politique démocratique)
 = b'. *Minu-neun minju-jeok-eulo jeongchi-(leul + E) ha-nda*
 Minu-**Aux** démocratiquement politique-(**Acc + E**) faire-**St**
 (Minu fait de la politique démocratiquement)

Nous observons que les phrases (46b) et (46b') sont sémantiquement quasi-synonymes. Nous pouvons écrire la relation de paraphrase entre la construction adjectivale et la construction adverbiale des noms figés en permettant la substitution de *-jeok(in)* par *-jeok-eulo* :

$$\begin{array}{llll} N_0 & Rx-(E + jeokin + jeok) & Npred-Acc & hada \\ = & N_0 & Rx-jeok-eulo & Npred-(Acc + E) \quad hada \end{array}$$

7-4-2. Type B : Rx (*-jeokin+ -han + E) Npred

Considérons le cas de *Rx-han*. Les phénomènes sont semblables à ceux de *Rx-jeok(in)* sauf que l'adjectif *Rx-hada* est combinable avec le suffixe adverbial *-ge*. Le morphème *-ge* s'applique régulièrement à tous les adjectifs²⁸ en *-hada*. Le suffixe *-ge* est un morphème de dérivation syntaxique d'adjectifs en adverbies.

(47) a. *ganglyeok daeeung (X Npred)(figé)*

(résolu - confrontation = confrontation résolue)

a'. *ganglyeok-ha-n daeeung (A-Sfd Npred)(libre)*

(résolu - être-Sfd - confrontation = confrontation résolue)

a''. *ganglyeok-ha-ge daeeung (Adv Npred)*

(résolument - confrontation = se faire face résolument)

b. *Minu-neun geu il-e ganglyeok-(han + E) daeeung-eul ha-yeot-da*

Minu-**Aux** ce affaire-**Postp.e** confrontation résolue-**Acc** faire-**Mpas-St**

(Minu a fait face résolument à cette affaire)

= b'. *Minu-neun geu il-e ganglyeok-ha-ge daeeung-(eul + E) ha-yeot-da*

Minu-**Aux** ce affaire-**Postp.e** résolument confrontation-(**Acc** + E) faire-**Mpas-St**

(Minu a fait face résolument à cette affaire)

Pour la structure *X Npred*, la relation dérivationnelle entre l'adjectif (*Rx-hada*) (47b) et l'adverbe (*Rx-hage*) (47b') met en jeu le suffixe adverbial *-ge* :

$$N_0 \quad Rx-(E+ han) \quad Npred-Acc \quad hada$$

²⁸ Les suffixes adverbiaux *-hi*, *-i*, *li*, *etc.* sont exclus dans ce travail, car ils produisent un nombre limité d'adverbes. C'est pourquoi tandis que l'on trouve des formes dérivées avec eux dans les dictionnaires éditoriaux, on n'y trouve pas les formes régulièrement dérivées avec le suffixe *-ge*.

= N_0 $Rx-ha-ge$ $Npred-(Acc + E)$ $hada$

Or, il est difficile de préciser si avec les séquences nominales $Rx-(jeok(in) + E)$ N ou $Rx-(han + E)$ N , la relation de transformation syntaxique mentionnée ci-dessus s'accompagne d'une modification du sens. Notre réponse à cette question est plutôt négative en général. Il nous semble que la différence sémantique entre les $Rx-jeokin$ N (ou $Rx-han$ N) et les X N est imperceptible, si et seulement si les N ont un emploi prédicatif dans la phrase.

Cependant, il existe des cas de X N figés à un très haut degré, par exemple, des termes terminologiques ou officiellement institutionnalisés. Par exemple, *yeonhap* (association) dans (48) est, en général, un nom prédicatif à verbe support *-hada*. Mais il a perdu la fonction prédicative dans (48a). La relation de paraphrase modifie le sens figé de cette séquence et lui fait perdre le statut de nom propre figé.

(48) a. *gukje-(E + *jeokin + *jeok)* *yeonhap* (XN)

(international - association = Organisation des Nations Unies)

b. *gukje-(jeokin + jeok)* *yeonhap* (($A-Sfd + N$) $Npred$)

(international ($A-Sfd + N$) – association = association qui a une envergure internationale ou association entre plusieurs pays)

8. Conclusion

Les sept informations fondamentales suivantes sont systématiquement décrites pour les 45000 noms figés à espacement facultatif dans notre dictionnaire électronique :

- (i) Espacement: ^ (espacement facultatif), blanc (espacement obligatoire)
- (ii) Informations flexionnelles: neuf classes de séquences nominales figées sont distinguées en fonction des contraintes combinatoires avec les postpositions nominales (N11, N12, N13, N21, N22, N23, N31, N32, N33). Chaque code appelle un transducteur de séquences de postpositions nominales.
- (iii) Catégorie grammaticale des éléments: N (nom), X (élément sans catégorie grammaticale), A (Adjectif), D (Déterminant), B (adverbe), etc.
- (iv) Origine : pk (proprement coréen), sk (sino-coréen), bor (autres emprunts), Mix (forme mélangée)
- (v) Informations sémantiques : Hum (nom humain), CollHum (nom humain collectif), InstHum (nom humain collectif institutionnel), Plt (nom de plante), Anm (nom d'animal), Inanim (nom inanimé).

(vi) Informations syntaxiques :

- si N_1N_2 être un N_1 et N_1N_2 être un N_2 , le code est +12
- si N_1N_2 être un N_1 mais $*N_1N_2$ être un N_2 , le code est +1
- si $*N_1N_2$ être un N_1 mais N_1N_2 être un N_2 , le code est +2
- si $*N_1N_2$ être un N_1 et $*N_1N_2$ être un N_2 , le code est +0
- si XN être un N , le code est +2
- si $*XN$ être un N , le code est +0
- si $*X_1X_2$ être un X_2 , le code est +0
- si NX être un N , le code est +1

(vii) Informations morpho-syntaxiques: -PRED (nom non prédicatif), +PRED (nom prédicatif). Les noms prédicatifs se subdivisent en trois catégories :

- « +PRED » indique un nom prédicatif, non associé à un verbe.
- « +PRED = V-n » indique un substantif entier morpho-syntaxiquement associé à un verbe
- « +PREDHN = V-n » indique un substantif dont le nom tête est morpho-syntaxiquement associé à un verbe

Pour 45000 *NFF*, nous comptons 68% de « -PRED », 24% de « +PREDHN = V-n », 7% de « +PRED » et 1% de « +PRED = V-n ».

Enfin, nous présentons des perspectives sur la description systématique pour chaque classe (*NC*, *NFF* et *NFO*) de séquences nominales figées les propriétés syntaxiques, sémantiques et flexionnelles dans un dictionnaire électronique. Voici un exemple de chaque classe de séquences nominales figées.

(49) a. *jeongchiga*, N11+sk-PRED+Hum//정치가(政治家)

a'. *kkotgalu*, N33+pk-PRED

b. *gukchaek^eunhang*, N12+NN+sk_sk-PRED+2+InstHum//국책은행(國策銀行)

c. *bingsanui ilgak*, N13+NuiN+sk_pk-sk+PRED+0//빙산(冰山)의 일각(一角)

Le nom dérivé *jeongchi-ga* « politique - personne = politicien » (49a) et le nom composé au sens traditionnel *kkot-galu* « fleur - poudre = pollen » (49a') sont des exemples de *NC*. *gukchaek ^ eunhang* « politique nationale - banque = banque

nationale » (49b) est un exemple de *NFF* et *bingsanui ilgak* « iceberg-Gén - coin = la partie émergée de l'iceberg » (49c) un exemple de *NFO*. Nous n'avons pas évoqué les formes écrites chinoises d'origine pour les éléments **sk** ou **bor** après « // », qui doivent être ajoutées dans notre dictionnaire.

Chapitre 4. Dictionnaires représentés par des graphes : séquences nominales figées et séquences de postpositions nominales

Dans le chapitre 3, nous avons discuté la représentation des séquences nominales figées sous forme de liste. Les informations mentionnées dans le chapitre 3 peuvent être décrites non seulement sous forme de liste mais aussi sous forme de graphes à l'aide d'INTEX. Dans ce chapitre, nous proposons de représenter d'autres types de séquences nominales et les postpositions nominales sous forme de graphes. Tout d'abord, la section 1 sera consacrée au choix du type de représentation et aux avantages de la représentation par des graphes. Dans les sections 2 et 3, nous présentons les dictionnaires électroniques représentés par des graphes : celui des séquences nominales figées (section 2) et celui des séquences de postpositions nominales (section 3).

1. Choix entre la représentation sous forme de liste ou de graphes

Dans le traitement des groupes nominaux en coréen, nous sommes confrontés à trois sortes d'agglutinations : entre noms, entre postpositions nominales, et entre noms et postpositions nominales. Par exemple :

- (1) a. *gongsang ^ gwahak ^ yeonghwaeseojochado* $N \wedge N \wedge N\text{-Postp-Postp-Postp}$
 [*gongsang gwahak yeonghwa*]-[*eseo-jocha-do*]
 [imagination science film]-[dans-même-aussi]
 = même dans les films de science-fiction

La séquence (1a) peut se réaliser graphiquement de quatre façons :

- b. *gongsang gwahak yeonghwaeseojochado* [*N N N*]-[*Postp Postp Postp*]
 gongsang gwahakyeonghwaeseojochado [*N NN*]-[*Postp Postp Postp*]
 gongsanggwahak yeonghwaeseojochado [*NN N*]-[*Postp Postp Postp*]
 gongsanggwahakyeonghwaeseojochado [*NNN*]-[*Postp Postp Postp*]

Dans l'exemple (1b), nous constatons que l'espacement entre noms est irrégulier, et que l'on ne laisse pas de blanc entre postpositions nominales ni entre noms et postpositions nominales.

D'une part, pour la reconnaissance automatique des séquences nominales figées

à espacement irrégulier, nous avons construit un dictionnaire électronique qui précise leurs variations typographiques. D'autre part, pour l'analyse automatique des séquences de postpositions nominales soudées, il est indispensable de disposer d'un dictionnaire électronique des séquences de postpositions nominales qui précise la segmentation, par exemple, à l'aide d'un transducteur fini. Enfin, à partir de ces deux dictionnaires, nous pouvons construire automatiquement le dictionnaire électronique des formes fléchies des séquences nominales figées¹ qui permet d'identifier les séquences nominales figées dans des textes.

Pour l'efficacité de la construction des dictionnaires électroniques de séquences nominales figées et de postpositions nominales, nous proposons deux types de représentation :

- la représentation sous la forme de liste pour les séquences nominales figées
- la représentation sous la forme de graphes pour certains autres types de séquences nominales et les postpositions nominales

Dans la représentation du lexique coréen par des graphes dans INTEX, nous observons au moins trois avantages :

- (i) Elle évite des répétitions : le graphe est plus compact que la liste et il est plus lisible dans le sens où il met en évidence les factorisations
- (ii) Elle explicite la relation d'équivalence qui lie les formes, elle permet donc de choisir un représentant unique et ses variantes (ou ses étiquettes linguistiques ou ses traductions) à l'aide de transducteurs
- (iii) Elle permet de décrire facilement par l'opérateur d'effacement « L » les modifications graphiques liées à la concaténation des postpositions nominales en coréen

Pour profiter de ces avantages, nous proposons donc de représenter sous la forme de graphes quelques autres types de séquences nominales assez factorisables dans la section 2 et les postpositions nominales dans la section 3.

2. Dictionnaires de séquences nominales représentées par des graphes

¹ cf. Le dictionnaire électronique de formes fléchies des séquences nominales figées, voir le chapitre 5.

2-1. Un exemple de graphe de séquences nominales à espacement facultatif

Nous pouvons décrire une série de séquences nominales à espacement facultatif à l'aide de l'éditeur graphique d'INTEX. Si les séquences nominales présentent des possibilités de factorisation, cette méthode de description permet d'éviter les répétitions.

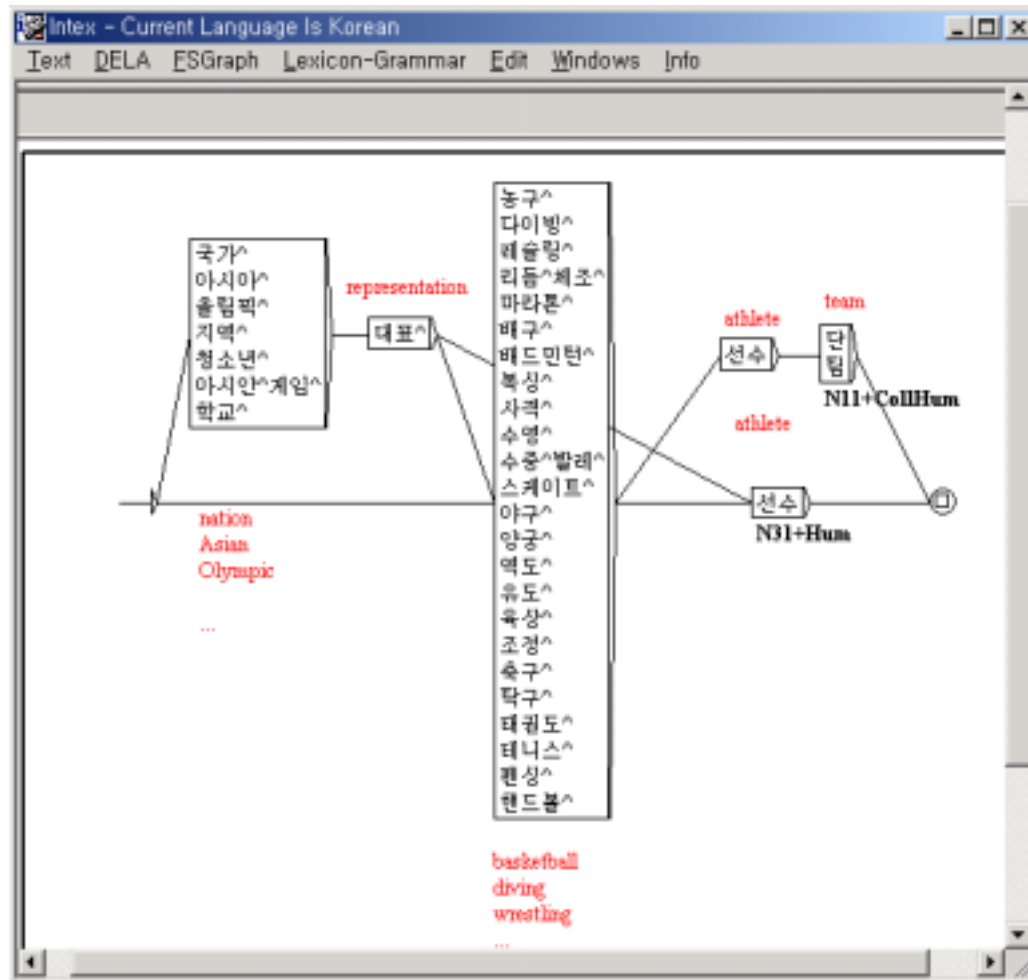


Figure 1. Graphe de noms figés à espacement facultatif

Extrayons de la figure 1 quelques exemples de chemins :

nonggu ^ seonsu,N31+Hum (= basketteur)

nonggu ^ seonsutim,N11+CollHum (= équipe de basket)

olimpik ^ daepyo ^ nonggu ^ seonsu,N31+Hum

(= basketteur sélectionné pour les jeux olympiques)

olimpik ^ daepyo ^ nonggu ^ seonsutim,N11+CollHum
 (= équipe olympique de basket)

À partir de ce simple graphe, une liste de 583 séquences nominales figées est générée par le module « Generate Language equivalent to FST » d'INTEX (figure 2).

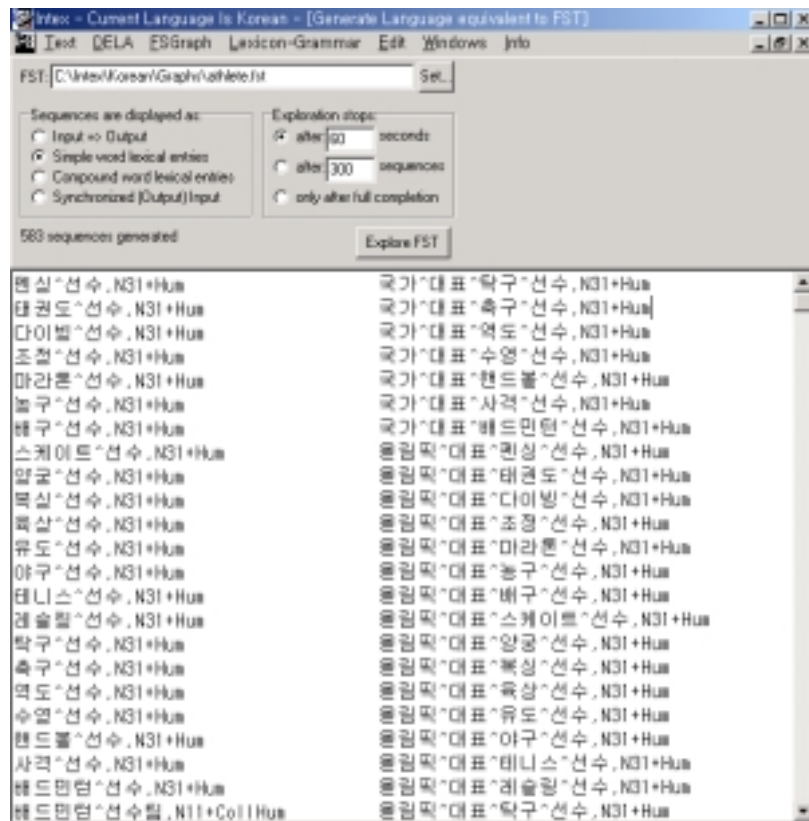


Figure 2. Génération d'une liste à partir d'un graphe

2-2. Un exemple de graphe de séquences nominales compactes et à espacement facultatif

L'exemple suivant est un graphe contenant des séquences nominales compactes et figées à espacement facultatif. Si nous voulons construire le dictionnaire des séquences nominales par des graphes, nous remarquons qu'il peut être intéressant de décrire à la fois des noms compacts et figés à espacement facultatif dans un même graphe. Quand un mot est figé avec un élément monosyllabique, ils ne sont séparés en général par aucun espace. En revanche, lorsqu'il est figé avec un élément à plus de deux syllabes, l'espacement est souvent facultatif.

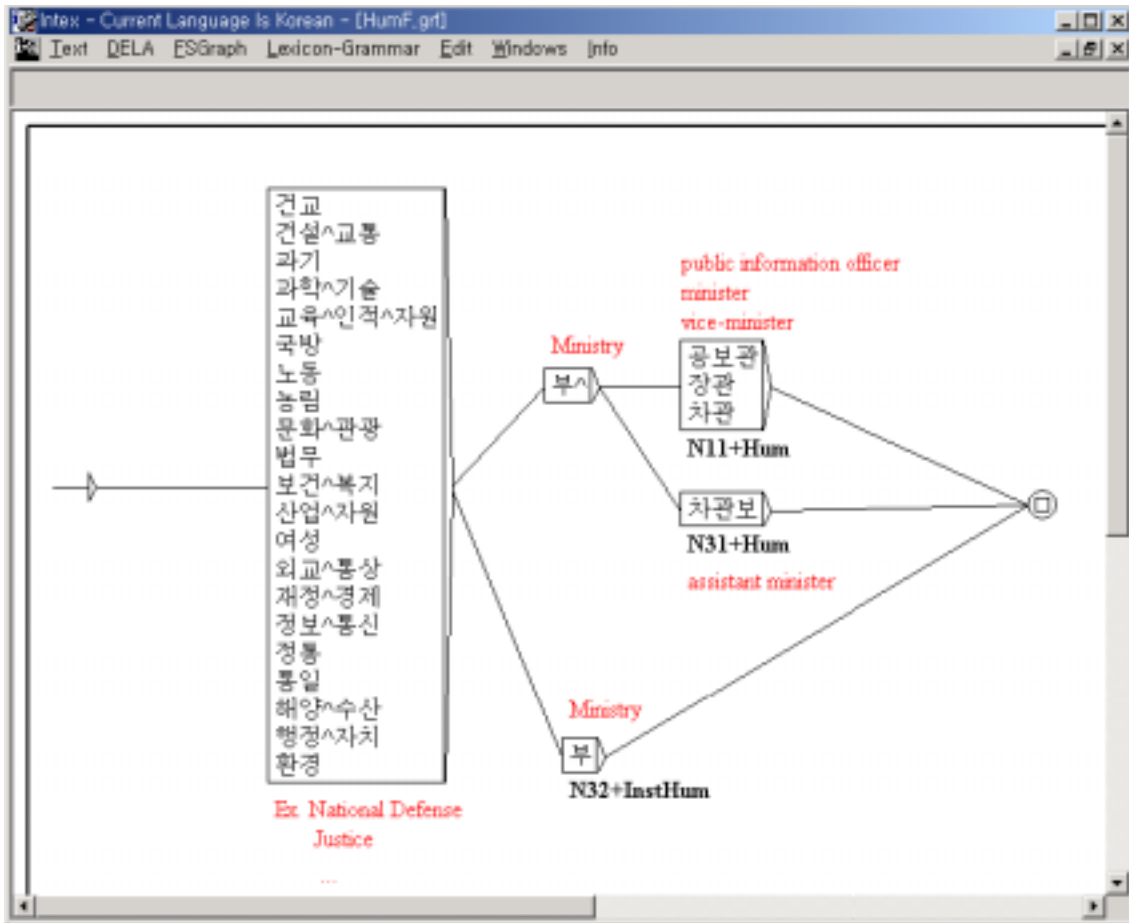


Figure 3. Graphe de noms figés à espacement facultatif et de noms compacts

Extrayons de la figure 3 quelques exemples de chemins :

nodongbu, N31+InstHum (= ministère du Travail)
tonilbu, N31+InstHum (= ministère de l'Unification)
beopmubu, N31+InstHum (= ministère de la Justice)
gwahak ^ gisulbu, N31+InstHum (= ministère des Sciences et Techniques)
nodongbu^janggwan, N11+Hum (= ministre du Travail)
tonilbu^janggwan, N11+Hum (= ministre de l'Unification)
beopmubu^janggwan, N11+Hum (= ministre de la Justice)
gwahak ^ gisulbu^janggwan, N11+Hum
 (= ministre des Sciences et Techniques)

À partir de ce simple graphe, une liste de 105 séquences nominales figées est générée par le module « Generate Language equivalent to FST » d'INTEX (figure 4).

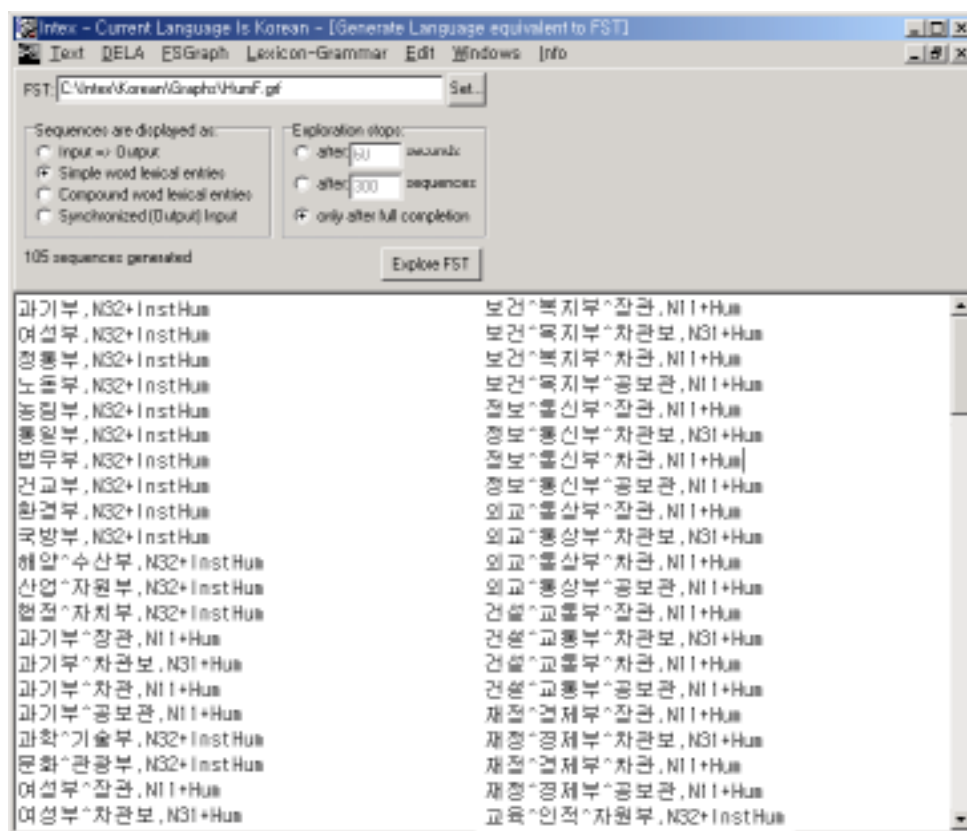


Figure 4. Génération d'une liste à partir du graphe de la figure 3

3. Dictionnaire de séquences de postpositions nominales représentées par des graphes

3-1. Classification linguistique des postpositions nominales

Avant de parler du dictionnaire des séquences de postpositions nominales, nous introduisons la classification linguistique des postpositions nominales. En coréen, la postposition nominale est une des parties du discours. Les postpositions sont classées en trois grandes catégories en fonction de critères syntaxiques et distributionnels, à savoir les postpositions casuelles, auxiliaires et conjonctives². Ces deux types de critères ont des relations étroites.

3-1-1. Critère syntaxique

² cf. Kim S.D (1992), Nam G.S et Go Y.G (1998), Lee H.J et Lee J.H (1998).

Dans la phrase coréenne, le déplacement des constituants est relativement libre par rapport au français ou à l'anglais. Les postpositions casuelles marquent le rôle syntaxique des substantifs. Ainsi, la postposition nominative *-ga/-i* attribue le cas nominatif aux substantifs. Les actants, y compris le sujet et l'objet direct, sont donc distinguables grâce aux postpositions³. Les postpositions casuelles attachées aux noms sont sous-catégorisées en six cas : le cas nominatif, le cas accusatif, le cas copulatif, le cas vocatif, le cas génitif, et une catégorie résiduelle qualifiée d'adverbiale ((2a) - (2f)).

- (2) a. *Minu-ga* *hakgyo-e* *ga-nda*
 Minu-**Nmtf** école-**Postp.e** aller-**St**
 (Minu va à l'école)
- b. *Minu-ga* *Ina-leul* *salangha-nda*
 Minu-**Nmtf** Ina-**Acc** aimer-**St**
 (Minu aime Ina)
- c. *geugeot-eun* *jongi-ga* *ani-da*
 Ce-**Aux** papier-**Copu** (ne pas être)-**St**
 (Ce n'est pas du papier)
- c'. *geugeot-eun* *jongi-i-da*
 Ce-**Aux** papier-être-**St**
 (C'est un papier)
- d. *Minu-ya,* *ppali* *o-neola*
 Minu-**Voc** vite venir-**StImp**
 (Minu, viens vite)
- e. *Minu-ui* *chaek-eul* *da-o*
 Minu-**Gén** livre-**Acc** donner-**StImp**
 (Donne-moi le livre de Minu)
- f. *Minu-ga* *Ina-ege* *kkot-eul* *ju-eot-da*
 Minu-**Nmtf** Ina-**Postp.ege** fleurs-**Acc** donner-**Mpas-St**
 (Minu a donné des fleurs à Ina)

Le nom propre *Minu* de (2a) est marqué comme sujet dans cette phrase par le cas

³ a. *Minu-ga* *Ina-leul* *salangha-nda*
 Minu-**Nmtf** Ina-**Acc** aimer-**St**
 (Minu aime Ina)

b. *Ina-leul* *Minu-ga* *salangha-nda*
 Ina-**Acc** Minu-**Nmtf** aimer-**St**
 (Minu aime Ina)

nominatif *-ga*. *Ina* dans (2b) est l'objet direct dans cette phrase en raison du cas accusatif *-leul*. La postposition *-ga/-i* suivie du verbe *anida* « ne pas être » marque le cas copulatif (2c), non le nominatif. Ici, nous distinguons le copulatif du nominatif. (2c) n'est qu'une forme négative de la phrase (2c') à copule *ida* « être »⁴. Dans (2c) et (2c'), *jongi* « papier » ne peut jamais être un sujet. Il ne joue que le rôle du complément de la copule *ida* « être » ou de sa forme négative *anida* « ne pas être » et enfin, le rôle de prédicat avec la copule. Si une postposition nominale ne joue jamais le rôle du sujet dans la phrase, on ne peut pas l'étiqueter comme nominatif. La postposition vocative de (2d) *-ya* donne la valeur d'appel à *Minu*. La postposition du génitif dans (2e) fait de *Minu* un modifieur du substantif *chaek* « livre ». Dans l'exemple (2f), *Ina* devient le datif par la postposition *-ege* « à ». *Ina* est un complément en *-ege*. L'instrumental (*-losseo/-eulosseo*), le locatif (*-e, -eseo*), le comparatif (*-boda, -mankeum*) etc. sont d'autres exemples de postpositions dites adverbiales. Nous utiliserons les notations *Postp.losseo, Postp.e, Postp.ege*, etc. pour ce type de postposition⁵.

Les postpositions auxiliaires ajoutent des nuances sémantiques à la phrase. Elles n'ont pas de rôle syntaxique direct. La fonction syntaxique des substantifs auxquels elles s'attachent n'est pas déductible des postpositions auxiliaires. Par exemple, la postposition auxiliaire *-neun* a la même forme dans les différentes positions syntaxiques. Voici quelques exemples :

- (3) a. *Ina-neun hakgyo-e ga-t-da*
 Ina-**Aux** école-**Postp.e** aller-**Mpas-St**
 (C'est Ina qui est allée à l'école)
- b. *Minu-ga Ina-neun salangha-nda*
 Minu-**Nmtf** Ina-**Aux** aimer-**St**
 (C'est Ina que Minu aime)
- c. *Minu-ga Ina-neun kkot-eul ju-eot-da*

⁴ Il y a plusieurs points de vue sur la copule *-ida* dans la linguistique coréenne. Comme en (2c'), lorsque les noms se combinent avec la copule *-ida*, ils ne demandent aucune postposition nominale. C'est pourquoi dans la grammaire scolaire, on considère *-ida* lui-même comme une postposition nominale et on l'appelle la postposition du prédicatif. Certaines linguistes apprécient *-ida* comme copule et considèrent les noms associés à *-ida* comme des noms prédicatifs. Mais tous les noms du coréen, qui sont combinables avec *-ida*, devraient donc être considérés comme des noms prédicatifs. À notre avis, les noms concrets comme *jongi* (papier) n'ont aucune propriété inhérente de prédicat (cf. pour la définition sur les noms prédicatifs, voir la section 7 du chapitre 3).

⁵ En grammaire traditionnelle, on a utilisé le terme *postposition adverbiale* pour ce type de postposition. Mais contrairement à ce que leur nom *postposition adverbiale* pourrait laisser penser, elles ne donnent pas nécessairement un caractère adverbial aux compléments où elles apparaissent.

Minu-**Nmtf** Ina-**Aux** fleurs-**Acc** donner-**Mpas-St**
 (C'est Ina à qui Minu a donné des fleurs)

Parmi les trois phrases précédentes, le nom propre *Ina* joue des rôles différents : dans (3a), le sujet ; dans (3b), le complément d'objet direct et dans (3c), le complément d'objet indirect. Les postpositions auxiliaires comme *-do* « aussi », *-neun* « c'est...que », *-jocha* « même », *-man* « ne...que », *-mada* « chaque », etc. apportent à la phrase des nuances sémantiques spécifiques⁶.

- (4) a. *Minu-ga* *Ina-man* *salangha-nda*
 Minu-**Nmtf** Ina-**Aux** aimer-**St**
 (Minu *n'*aime *que* Ina)
- b. *Minu-ga* *Ina-do* *salangha-nda*
 Minu-**Nmtf** Ina-**Aux** aimer-**St**
 (Minu aime *aussi* Ina)

Les postpositions conjonctives lient des substantifs ayant la même fonction syntaxique dans la phrase. Des substantifs liés par une postposition conjonctive sont considérés comme un constituant.

- (5) *Minu-wa* *Ina-ga* *hakgyo-e* *ga-nda*
 Minu-**Conj** Ina-**Nmtf** école-**Postp.e** aller-**St**
 (Minu et Ina vont à l'école)

Dans l'exemple ci-dessus la postposition de coordination *-wa* lie *Minu* et *Ina* qui devient une unité suivie de la postposition nominative *-ga*. Les postpositions conjonctives ressemblent à des postpositions casuelles par plusieurs aspects. Mais elles ne peuvent pas attribuer de fonction syntaxique aux substantifs et seulement les coordonner.

3-1-2. Critères distributionnels

Certaines postpositions auxiliaires peuvent s'attacher à des adverbes pour

⁶ En revanche, la postposition de cas nominatif, *-ga* de (2a) et la postposition de cas accusatif, *-leul* de (2b) ne portent pas de sens. Le substantif accolé avec le cas nominatif *-ga* est le sujet et *-leul*, l'objet direct. La postposition conjonctive *-wa* « et » de (5) n'a que le sens de coordination.

ajouter une nuance sémantique. Les postpositions casuelles et conjonctives n'ont pas cette propriété.

- (6) a. *Ina-ga jal-do geot-neunda*
 Ina-**Nmtf** bien-**Aux** marcher-**St**
 (Ina marche bien *aussi*)
- b. *Ina-ga jal-man geot-neunda*
 Ina-**Nmtf** bien-**Aux** marcher-**St**
 (Ina ne marche que bien)
- c. **Ina-neun jal (-ga, -eul, -ui, -lo, -ya, -wa) geot-neunda*
 *Ina-**Aux** bien (**Nmtf, Acc, Gén, Postp.lo, Voc, Conj**) marcher-**St**

Les postpositions casuelles et conjonctives peuvent être omises dans la langue orale si et seulement si les fonctions syntaxiques des substantifs sans postpositions sont suffisamment claires. En particulier, les cas nominatif et accusatif sont aisément omis. En revanche, les postpositions auxiliaires ne peuvent pas être omises sans changement de sens. Par exemple :

- (7) a. *geu namja-(ga + E) tteona-t-ni ?*
 Cet homme-(**Nmtf** + E) partir-**Mpas-Sti** ?
 (Cet homme, est-il parti ?)
- b. *na-do i chaek -(eul + E) il-eot-e*
 Je-Aux ce livre-(**Acc** + E) lire-**Mpas-St**
 (Moi aussi, j'ai lu ce livre)
- c. *hakgyo-(e + E) ga-ni ?*
 école-(**Postp.e** + E) aller-**Sti** ?
 (Est-ce que tu vas à l'école ?)
- d. *donggul-(ui + E) ipgu*
 grotte (**Gén** + E) entrée
 (L'entrée de la grotte)
- e. *chaek (-gwa + E,) yeonpil-eul da-o*
 livre -(**Conj** + E,) stylo-**Acc** donner-**StImp**
 (Donnez-moi un livre et un stylo)

Nous pouvons résumer les critères de classification des postpositions nominales dans la figure 5.

		Postp. Casuelle	Postp. Auxiliaire	Postp. Conjonctive
Critère syntaxique	Marque la fonction syntaxique des actants	Oui	Non	Non
Critères distributionnels	Possibilité de combinaison avec des adverbes	Non	Oui	Non
	Omission	Oui	Non	Oui

Figure 5. Les critères de classification des postpositions nominales

3-2. Le marqueur pluriel *-deul* en coréen

3-2-1. Problématique du marqueur pluriel *-deul*

Avant de décrire des graphes de postpositions nominales, nous présentons ici le comportement spécifique du morphème pluriel *-deul* en coréen. Les marques de pluriel dans les langues indo-européennes occupent une place importante. On observe toujours l'accord en nombre entre le sujet et le verbe ou bien entre le substantif déterminé et son déterminant :

- (8) a. *Mon ami est allé au cinéma.*
 b. *Mes amis **sont** allés au cinéma.*

Le coréen a un seul marqueur pluriel *-deul* qui est comparable au morphème pluriel *-s* en anglais, et on n'accorde pas en nombre le substantif en position sujet et le verbe ni le substantif déterminé et son déterminant. Bien que le verbe et l'adjectif en coréen se conjuguent avec des suffixes verbaux ou adjectivaux, il n'existe aucun suffixe de conjugaison obligatoire qui morphologiquement marque le pluriel du sujet. Dans (9a) et (9b) les formes conjuguées des verbes sont identiques et indépendantes du nombre des sujets.

- (9) a. *nae chingu-ga geukjang-e ga-t-da*
 Mon ami-**Nmtf** cinéma-**Postp.e** aller-**Mpas-St**
 (Mon ami est allé au cinéma)

- b. *nae chingu-deul-i geukjang-e ga-t-da*
 Mon ami-**Mplu-Nmtf** cinéma-**Postp.e** aller-**Mpas-St**
 (Mes amis sont allés au cinéma)

Par ailleurs, même en combinaison avec des noms comptables, *-deul* est souvent facultatif. Sans contexte, le nom comptable sans *-deul* est souvent ambigu quant au nombre :

- (10) a. *i seojeom-eseo Minu-ga chaek-eul sa-t-da*
 ce librairie-**Postp.eseo** Minu-**Nmtf** livre-**Acc** acheter-**Mpas-St**
 (Dans cette librairie Minu a acheté (un + des) livre(s))
 b. *i seojeom-eseo Minu-ga chaek-deul-eul sa-t-da*
 ce librairie-**Postp.eseo** Minu-**Nmtf** livre-**Mplu-Acc** acheter-**Mpas-St**
 (Dans cette librairie Minu a acheté des livres)
 c. *seojeom-e-neun chaek-i it-da*
 librairie-**Postp.e-Aux** livre-**Nmtf** il y a-**St**
 (Dans la librairie il y a des livres)
 d. *i seojeom-e-neun chaek-(deul + E)-i mani it-da*
 ce librairie-**Postp.e-Aux** livre-(**Mplu+ E**)-**Nmtf** beaucoup il y a-**St**
 (Dans cette librairie il y a beaucoup de livres)

Dans (10a), nous ne pouvons pas savoir si *Minu* a acheté un ou plusieurs livres sans un contexte spécifique. En coréen, le marqueur de pluriel apparaît facultativement. Mais surtout pour les noms inanimés comme *chaek* « livre », (10a) sans *-deul* est beaucoup plus naturelle que (10b), même si le sens est le pluriel. En (10c), nous devinons le pluriel par le contexte ou pour des raisons extra-linguistiques (en général, dans une librairie il y a beaucoup de livres). Un déterminant ou un adverbe de quantité comme *mani* « beaucoup », *modu* « tous », *jeonbu* « tous », *jogeum* « un peu », etc. remplace souvent *-deul* comme en (10d), ce qui rend la phrase plus naturelle.

Cependant, le nom inanimé *chaek* « livre » doit être obligatoirement pluralisé dans un contexte anaphorique⁷. Un nom anaphorisant (11a') doit concorder pour le nombre avec le nom anaphorisé (11a) :

- (11) a. *Minu-ga chaek se gwon-eul sa-t-da*

⁷ cf. Pour des exemples dans des contextes anaphoriques, voir Jun J.Y (1999), pp. 193-194.

Minu-**Nmtf** livre trois volume [Nclass]-**Acc** acheter-**Mpas-St**
 (Minu a acheté trois volumes de livres)

a' Minu-ga **geu** chaek(*E + **-deul**)-eul ilg-eot-da
 Minu-**Nmtf** ce livre(*E +-**Mplu**)-**Acc** lire-**Mpas-St**
 (Minu a lu (*ce livre + ces livres))

Avec les déterminants démonstratifs comme *jeo* ou *i* (en français *ce*), il faut obligatoirement préciser le nombre du nom. L'absence de *-deul* implique le singulier :

- (12) a. *i seojeom-eseo* Minu-ga *jeo* chaek-eul sa-t-da
 ce librairie-**Postp.eseo** Minu-**Nmtf** ce livre-**Acc** acheter-**Mpas-St**
 (Dans cette librairie Minu a acheté ce livre)
- b. *i seojeom-eseo* Minu-ga *jeo* chaek-deul-eul sa-t-da
 ce librairie-**Postp.eseo** Minu-**Nmtf** ce livre-Mplu-**Acc** acheter-**Mpas-St**
 (Dans cette librairie Minu a acheté ces livres)

Pour les noms animés comptables, la forme sans *-deul* exprime en général le sens singulier :

- (13) a. *na-neun chingu-hago* jeomsim-eul meok-eot-da
 Je-**Aux** ami-**Postp.hago** déjeuner-**Acc** manger-**Mpas-St**
 (J'ai déjeuné avec mon ami)
- b. *na-neun chingu-deul-hago* jeomsim-eul meok-eot-da
 Je-**Aux** ami-**Mplu-Postp.hago** déjeuner-**Acc** manger-**Mpas-St**
 (J'ai déjeuné avec mes amis)

Mais dans certains contextes où le référent d'un nom animé comptable n'est pas concret ni ancré dans le temps et dans l'espace, la forme sans *-deul* n'exclut pas l'interprétation au sens pluriel comme en (14a) :

- (14) a. *hakgyo-eseo-neun chingu-hago* jal jinae-yaha-nda
 école-**Postp.eseo-Aux** ami-**Postp.hago** bien vivre-devoir-**St**
 (Au moins dans l'école, on doit s'entendre bien avec (son ami + ses amis))
- b. *hakgyo-eseo-neun chingu-deul-hago* jal jinae-yaha-nda
 école-**Postp.eseo-Aux** ami-**Mplu-Postp.hago** bien vivre-devoir-**St**
 (Au moins dans l'école, on doit s'entendre bien avec ses amis)

Comme nous l'avons examiné ci-dessus, les interprétations des formes sans -*deul* ne sont pas toujours claires quant au nombre sans un contexte précis. Quant aux positions de -*deul* dans un groupe nominal, il se place après les séquences nominales et avant les autres postpositions nominales. Par exemple :

- (15) a. *chaek-deul-(i + eul + ui + eun + lo)*
 livre-**Mplu**-(**Nmtf** + **Acc** + **Gén** + **Aux** + **Postp.lo**)
 b. **chaek-(i + eul + ui + eun)-deul*
 *livre-(**Nmtf** + **Acc** + **Gén** + **Aux**)-**Mplu**

Cependant, il se place parfois après des postpositions nominales dites adverbiales⁸ :

- c. *chaek-(eulo + e)-deul*
 livre-(**Postp.lo** + **Postp.e**)-**Mplu**

Or, l'emploi de -*deul* après des postpositions nominales dites adverbiales dans (15c) est différent de celui dans (15a).

En coréen, le marqueur pluriel -*deul* a plusieurs possibilités de placement. Lorsqu'il se combine avec un substantif comptable, il marque le pluriel de ce substantif lui-même comme en (16a). Si ce morphème se combine avec un substantif non comptable ou avec d'autres éléments comme un adverbe, les postpositions nominales dites adverbiales, etc., il indique le pluriel du sujet de la phrase comme en (16b) et (16c).

- (16) a. *Minu-ga chaek se gwon-eul sa-t-da*
 Minu-**Nmtf** livre trois volume-**Acc** acheter-**Mpas-St**
 (Minu a acheté trois volumes de livres)
*Minu-ga geu chaek(*E + -deul)-eul ilg-eot-da*
 Minu-**Nmtf** ce livre(*E +-**Mplu**)-**Acc** lire-**Mpas-St**
 (Minu a lu (*ce livre + ces livres))

⁸ Selon les intuitions des linguistes, le jugement d'acceptabilité de l'exemple suivant est discutable. Tandis que Kuh H. (1987) accepte de tels exemples, Jun J.Y (1999) ne les accepte pas. Nous ne pouvons pas répondre immédiatement à cette question.

? *i ai-deul-hago-deul jal jinae-la*
 ce enfant-**Mplu-Postp.hago-Mplu** bien vivre-**StImp**
 (Entendez-vous bien avec ces enfants) (Jun, J.Y, 1999)

- b. (*neohui-deul* + *E*), *ppali-deul* *o-neola*
 vite-**Mplu** venir-**StImp**
 (Venez vite)
- c. (*neohui-deul* + *E*), *bap-deul* *meok-go* *it-ni?*
 riz-**Mplu** manger-**Sconj** il y a-**Sti?**
 (Est-ce que vous êtes en train de manger du riz ?)

Ainsi, la distribution du marqueur pluriel *-deul* est plus variable que celle du suffixe pluriel *-s* en français ou en anglais et ses conditions d'emploi sont complexes. Dans la section 3-2-2, nous examinerons trois cas de restrictions distributionnelles sur *-deul*, et dans la section 3-2-3, les deux emplois de *-deul* : le marqueur de pluriel du substantif et le marqueur de pluriel du sujet de la phrase. Dans la section 3-2-4, nous constaterons la relation de paraphrase entre les phrases dans lesquelles *-deul* marque le pluriel du sujet en plusieurs positions.

3-2-2. Restrictions distributionnelles sur *-deul*⁹

Cas 1. N_0 = : nom non comptable ($*N_0$ -*deul* *W* *V*)

-deul ne s'emploie pas avec le nom sujet non comptable.

- (17) a. *mul*-(*E* + $*deul$)-*i* *mani* *heuleu-nda*
 eau-**Nmtf** beaucoup couler-**St**
 (Beaucoup d'(eau + $*eaux$) (coule + $*coulent$))
- b. *gongbu*-(*E* + $*deul$)-*ga* *swip-da*
 etude-(*E* + $*Mplu$)-**Nmtf** être facile-**St**
 ((L'étude + $*Les$ études) (est + $*sont$) facile)

Cas 2. N_i = : numéral ou nom classifieur ($*N_i$ -*deul*)

L'apparition des noms numéraux (*Nnum*) comme *set* « trois » ou des noms classifieurs (*Nclass*) comme *myeong* « personne » bloque l'occurrence de *-deul* en n'importe quelle position.

⁹ cf. Nam G.S et Go Y.G (1998), Seo J. S (1996). En outre, Jun J.Y (1999) a remarqué que certains noms collectifs comme *illyu* (espèce) ne se combinent pas avec *-deul*.

- (18) a. *na-neun haksaeng set-(E+ *deul)-eul mana-t-da*
 Je-**Aux** étudiant trois-(E+ ***Mplu**)-**Acc** rencontrer-**Mpas-St**
 (J'ai rencontré les trois étudiants)
- = b. *na-neun haksaeng se myeong-(E+ *deul)-eul mana-t-da*
 Je-**Aux** étudiant trois personne-(E+ ***Mplu**)-**Acc** rencontrer-**Mpas-St**
 (J'ai rencontré les trois étudiants)

Cas 3. $N_i =$: nom dans un emploi générique ($*N_i-deul$)¹⁰

Dans l'emploi générique d'un nom comptable, la forme du nom avec *-deul* est exclue.

- (19) a. *salam-eun manmul-ui yeongjang-i-da*
 homme-**Aux** univers-**Gén** roi-**être-St**
 (L'homme est le roi de la création)
- b. **salam-deul-eun manmul-ui yeongjang-i-da*
 homme-**Mplu-Aux** univers-**Gén** roi-**être-St**

3-2-3. Deux emplois de *-deul* : pluriel du substantif et pluriel du sujet

Dans la section précédente, nous avons vu que *-deul* ne peut suivre un nom sujet non comptable ($*N_0-deul$ W V). Nous allons voir comment *-deul* marque le pluriel (soit du substantif soit du sujet de la phrase) dans le cas où N_0 est un nom comptable.

Pour observer l'influence du pluriel de N_0 sur N_1 , nous examinons *-deul* dans les deux structures suivantes :

$$\begin{array}{lll} [N_0-Nmtf & N_1-Nmtf & V]^{11} \\ [N_0-Nmtf & N_1-Acc & V] \end{array}$$

dans lesquelles nous constaterons que si N_1 est un nom non comptable, N_1-deul marque le pluriel du sujet N_0 . Sinon, N_1-deul marque le pluriel de N_1 .

¹⁰ Pour l'emploi générique, la postposition auxiliaire *<-neun>* « c'est ... que », qui est un marqueur de la topicalisation dans une phrase, s'attache, en général, aux noms en position sujet.

¹¹ La construction $[N_0-Nmtf \quad N_1-Nmtf \quad V]$ est souvent appelée *phrase à double sujet*. Ici, les deux postpositions *-Nmtf* sont formellement identiques.

Cas 1. N_0 = : nom comptable et N_1 = : nom comptable

Regardons l'exemple (20) où N_0 est au singulier et N_1 est au singulier :

- (20) a. *nae chingu-ga jip-i it-da*
mon ami-**Nmtf** maison-**Nmtf** il y a-**St**
(Mon ami a (une + des) maison(s))
- b. *nae chingu-ga jip-eul sa-t-da*
mon ami-**Nmtf** maison-**Acc** acheter-**Mpas-St**
(Mon ami a acheté (une + des) maison(s))

Comme dans l'exemple (10a) (p. 113) où *-deul* n'apparaît pas, nous ne pouvons pas savoir si N_0 = : *chingu* « ami » a acheté une ou plusieurs maisons sans un contexte spécifique.

Dans (21), N_0 est au pluriel et N_1 est au singulier :

- (21) a. *nae chingu-deul-i jip-i it-da*
mon ami-**Mplu-Nmtf** maison-**Nmtf** il y a-**St**
(Mes amis ont (une + des) maison(s))
- b. *nae chingu-deul-i jip-eul sa-t-da*
mon ami-**Mplu-Nmtf** maison-**Acc** acheter-**Mpas-St**
(Mes amis ont acheté (une + des) maison(s))
- b'. *nae chingu-deul-i don-eul mo-a jip-eul sa-t-da*
mon ami-**Mplu-Nmtf** argent-**Acc** rassembler-**Sconj** maison-**Acc** acheter-**Mpas-St**
(Mes amis ont acheté une maison avec leur argent rassemblé)

Dans les exemples de (21), N_0 -*deul* est le pluriel de N_0 = : *chingu* « ami ». Mais nous ne pouvons pas non plus préciser le nombre de N_1 = : *jip* « maison » sans un contexte précis comme (21b'), parce qu'il se peut que mes amis soient les propriétaires d'une même maison ou de plusieurs maisons. Sans le marqueur *-deul*, l'interprétation de N_1 comptable est souvent ambiguë en nombre et le pluriel de N_0 ne donne pas nécessairement un sens pluriel à N_1 .

Maintenant, insérons *-deul* dans (21).

- (22) a. *nae chingu-deul-i jip-deul-i it-da*

mon ami-**Mplu-Nmtf** maison-**Mplu-Nmtf** il y a-**St**
(Mes amis ont des maisons)

b. *nae chingu-deul-i jip-deul-eul sa-t-da*
mon ami-**Mplu-Nmtf** maison-**Mplu-Acc** acheter-**Mpas-St**
(Mes amis ont acheté des maisons)

Dans (22a) - (22b), on ne sait pas si chaque ami a sa maison (ou a acheté sa maison) ou plusieurs maisons, mais au moins $N_I = : jip$ « maison » est pluriel. Remplaçons le sujet par le nom au singulier :

(23) a. *nae chingu-ga jip-deul-i it-da*
mon ami-**Nmtf** maison-**Mplu-Nmtf** il y a-**St**
(Mon ami a des maisons)

b. *nae chingu-ga jip-deul-eul sa-t-da*
mon ami-**Nmtf** maison-**Mplu-Acc** acheter-**Mpas-St**
(Mon ami a acheté des maisons)

Ici, *-deul* de $N_I = : maison$ « *jip* » marque le pluriel de N_I et le nombre de N_I n'a aucune relation avec celui de N_0 .

Nous pouvons donc conclure que si N_0 et N_I sont des noms comptables, *-deul* de N_0 marque le pluriel de N_0 et *-deul* de N_I celui de N_I . Mais avec la forme sans *-deul* en position N_I , sans contexte précis, nous ne pouvons pas déterminer si N_I est singulier ou pluriel. Enfin, le pluriel de N_I est indépendant du pluriel de N_0 . Récapitulons notre discussion sur le Cas 1 :

Cas 1. $N_0 = : \text{nom comptable}$ et $N_I = : \text{nom comptable}$				
(1) Si $N_0 = : S$	et	$N_I = : S$,	alors c'est $N_0 = : S$	et $N_I = : S/P (?)$
(2) Si $N_0 = : S$	et	$N_I = : P$,	alors c'est $N_0 = : S$	et $N_I = : P$
(3) Si $N_0 = : P$	et	$N_I = : S$,	alors c'est $N_0 = : P$	et $N_I = : S/P (?)$
(4) Si $N_0 = : P$	et	$N_I = : P$,	alors c'est $N_0 = : P$	et $N_I = : P$

Figure 6. Le nombre de N_I dans le cas 1¹².

¹² « S » indique le singulier et « P » pluriel.

Cas 2. N_0 = : nom comptable et N_1 = : nom non comptable

Jun J. Y (1999) distingue les propriétés inhérentes du nom (c'est-à-dire il s'agit de l'emploi générique dans (19), p. 117) de ses propriétés stéréotypiques :

- (24) a. *golae-(E + *deul)-neun* *poyudongmul-i-da*
 baleine-(E+ **Mplu**)-**Aux** mammifère-être-**St**
 (La baleine est un mammifère)
- b. *asiain-(E + deul)-eun* *meoli-ga* *geom-da*
 asiatique-(E + **Mplu**)-**Aux** cheveu-**Nmtf** être noire-**St**
 ((Un Asiatique + Les Asiatiques) (a + ont) les cheveux noirs)

(24a) est une phrase qui décrit une propriété inhérente, alors que (24b) est celle qui décrit une propriété non inhérente, mais stéréotypique. Dans (24b), l'occurrence de – *deul* est facultative.

Dans l'exemple suivant, N_0 : = *beol* « abeille » est un nom comptable et N_1 : = *kkul* « miel » est un nom non comptable. Le nom singulier comptable en position sujet de (25a) peut avoir un sens pluriel, lorsque la phrase décrit une propriété stéréotypique :

- (25) a. *beol-eun* *kkul-eul* *mandeu-nda*
 abeille-**Aux** miel-**Acc** produire-**St**
 (L'abeille produit du miel)
- = b. *beol-deul-eun* *kkul-eul* *mandeu-nda*
 abeille-**Mplu-Aux** miel-**Acc** produire-**St**
 (Les abeilles produisent du miel)

Maintenant, considérons les cas en question où N_1 est un nom non comptable en excluant les deux emplois générique et stéréotypique de N_0 .

- (26) a. *nae* *chingu-ga* *don-i* *eop-da*
 mon ami-**Nmtf** argent- **Nmtf** (ne pas avoir)-**St**
 (Mon ami n'a pas d'argent)
- b. **nae* *chingu-ga* *don-deul-i* *eop-da*
 *mon ami-**Nmtf** argent-**Mplu-Nmtf** (ne pas avoir)-**St**
- c. *nae* *chingu-deul-i* *don-i* *eop-da*
 mon ami-**Mplu-Nmtf** argent-**Nmtf** (ne pas avoir)-**St**

- (Mes amis n'ont pas d'argent)
- = d. *nae chingu-deul-i don-deul-i eop-da*
mon ami-**Mplu-Nmtf** argent-**Mplu-Nmtf** (ne pas avoir)-**St**
(Mes amis n'ont pas d'argent)
- (27) a. *nae chingu-ga geu gyeonggi-e don-eul geol-eot-da*
mon ami-**Nmtf** ce match-**Postp.e** argent-**Acc** parier-**Mpas-St**
(Mon ami a parié de l'argent à ces matches)
- b. **nai chingu-ga geu gyeonggi-e don-deul-eul geol-eot-da*
*mon ami-**Nmtf** ce match-**Postp.e** argent-**Mplu-Acc** parier-**Mpas-St**
- c. *nae chingu-deul-i geu gyeonggi-e don-eul geol-eot-da*
mon ami-**Mplu-Nmtf** ce match-**Postp.e** argent-**Acc** parier-**Mpas-St**
(Mes amis ont parié de l'argent à ces matches)
- = d. *nae chingu-deul-i geu gyeonggi-e don-deul-eul geol-eot-da*
mon ami-**Mplu-Nmtf** ce match-**Postp.e** argent-**Mplu-Acc** parier-**Mpas-St**
(Mes amis ont parié de l'argent à ces matches)

N_I : = *don* « de l'argent » est un nom non comptable. Si N_0 n'est pas au pluriel, N_I non comptable ne peut pas se combiner avec *-deul* comme (26b) et (27b). Le fait que N_0 soit au pluriel permet la combinaison entre *don* « de l'argent » et *-deul*. Les exemples (26c) et (26d), et (27c) et (27d) sont sémantiquement équivalents. Le marqueur de pluriel de N_I est facultatif et ne modifie pas le sens de la phrase. Ici, le pluriel du nom N_I non comptable ne marque que celui du sujet N_0 . On peut donc conclure que *-deul* avec le nom non comptable marque le pluriel du sujet de phrase. Nous pouvons résumer le nombre de N_I dans le Cas 2 comme suit :

Cas 2. N_0 = : nom comptable et N_I = : nom non comptable		
(1) Si N_0 = : S	et N_I = : S,	alors N_0 = : S
(2) N_0 = : S	et $*N_I$ = : P	: impossible
(3) Si N_0 = : P	et N_I = : S,	alors N_0 = : P
(4) Si N_0 = : P	et N_I = : P,	alors N_0 = : P

Figure 7. Le nombre de N_I dans le cas 2

Cas 3. ((N_0 -*deul*)+ *E*) *W-deul* *V* où N_0 = : nom comptable

Dans les cas 3 et 4, nous examinerons les conditions d'apparition de *-deul* en plusieurs positions dans la structure [N_0 *W* *V*] et constaterons que *-deul* en plusieurs positions marque le pluriel du sujet. Etant donné qu'il n'y a aucun marqueur de pluriel obligatoire dans les suffixes verbaux ou adjectivaux pour marquer le pluriel du sujet de la phrase, *-deul* joue ce rôle en se combinant avec plusieurs éléments. Cela produit des combinaisons entre *-deul* et N_1 non comptable, des adverbes, des postpositions nominales dites adverbiales, etc. Surtout, dans les phrases où le sujet est omis¹³, l'occurrence *-deul* s'observe fréquemment et elle marque le pluriel du sujet. Etant donné qu'il copie la pluralité du sujet en se combinant avec divers constituants du prédicat, ce *-deul* est appelé *marqueur pluriel copié* (MPC)(Plural Marker Copying, Song S. Ch, 1988)¹⁴ pour indiquer qu'il marque le pluriel du sujet dans les phrases.

Ici, en examinant plusieurs exemples, nous allons voir comment *-deul* peut se combiner avec des mots de diverses catégories grammaticales pour marquer le pluriel du sujet (N_0) dans la structure [N_0 *W* *V*] (cf. pour la relation de paraphrase, voir la section 3-2-4).

- (28) a. (*neohui-deul*¹⁵ + *E*), *ppali-deul* *o-neola*
 (vous + *E*) vite-**Mplu** venir-**StImp**
 (Venez vite)
- b. (*uli* + *E*), *gachi-deul* *ga-ja*¹⁶
 (nous + *E*) ensemble-**Mplu** aller-**StP**
 (Allons-y ensemble)

¹³ En coréen, le sujet est souvent omis dans les phrases interrogatives, propositives, impératives et même déclaratives.

¹⁴ Lee H. G (1991), Moon S. Ch (1995) et Jun J. Y (1999) adoptent ce terme PMC.

¹⁵ Les formes du pluriel des pronoms personnels *na* « je » et *neo* « tu » sont *uli* « nous » et *neohui* « vous ». En général, le marqueur pluriel *-deul* se combine facultativement avec ces formes modifiées : (**na* + *uli*)-*deul* et (**neo* + *neohui*)-*deul*.

a. (*uli* + *neohui*)-(E-*ga+deul-i*) *nolaeha-nda*
 (nous + vous)-(E-**Nmtf** + **Nmtf-Mplu**) chanter-**St**
 ((Nous + Vous) (chantons + chantez))

b. *neohui*-(*deul* + **E*), *ppali-deul* *o-neola*
 (vous + **E*) vite-**Mplu** venir-**StImp**
 (Venez vite)

Mais les formes plurielles (*neohui* et *uli*) des pronoms personnels sans *-deul* sont inacceptables dans certains contextes impératifs. En revanche, les pronoms personnels *geu* « il » et *jagi* « lui-même ou elle-même » ne sont pas modifiés au pluriel : (*geu* + *jagi*)-*deul*.

¹⁶ Les formes *-ja*, *-sita* (l'emploi honorifique) et *-sipsida* (l'emploi plus honorifique que *-sida*) sont des suffixes terminaux verbaux de mode propositif.

- (31) Q : (*neohui-deul* + E), *eodi-deul* *ga-ni* ?
 (vous + E), où-**Mplu** aller-**Sti** ?
 (Où allez-vous ?)
- R : (*uli-deul-eun* + E) *geukjang-(e + *deul)* *ga-yo*.
 (Nous-**Aux** +E) cinéma-(**Postp.e** + ***Mplu**) aller-**St**
 (Nous allons au cinéma) (Jun J. Y, 1999)

Cas 4. ((*N₀-deul*) + E) W V-*deul* où *N₀* = : nom comptable¹⁹

La combinaison entre *-deul* et les éléments de verbes ou d'adjectifs est également facile à observer. La fonction de *-deul* est ici aussi de marquer le pluriel du sujet. Il y a trois positions où *-deul* peut s'insérer : entre deux verbes, entre le radical sans catégorie grammaticale *Rx* et le suffixe *-hada* ou entre le nom prédicatif et le suffixe *-hada* et enfin, après le suffixe complémenteur.

(i) Entre deux verbes

- (32) (*yeoleobun-deul* + E), *dol-aseo-(deul + E)* *ga-seyo*
 (vous + E) retourner-**Sconj**-(**Mplu** + E) aller-**StImp**
 (Faites demi-tour et allez-y)

Le marqueur de pluriel *-deul* peut s'insérer entre deux verbes *dolda* « retourner » et *gada* « aller » coordonnés. Dans (32), *-aseo* joue le rôle d'un suffixe conjonctif coordonnant²⁰ et apporte le sens de succession (faire demi-tour et s'en aller). Ici, on ne doit pas confondre les deux cas de deux verbes coordonnés (32) et d'un verbe complexe (34) (p. 126). La différence entre les deux cas (32) et (34) est liée à la possibilité d'insertion de *-seo* (le suffixe conjonctif coordonnant de succession) : s'il n'y a qu'une coordination entre deux verbes et que ces deux verbes sont liés par l'action successive, l'insertion de *-seo* est permise²¹. En revanche, s'il y a une composition entre deux

¹⁹ V signifie soit une forme radicale *Rx* soit une forme combinée avec des suffixes verbaux ou adjectivaux.

²⁰ cf. Han S. H (2000, p.36) a distingué des suffixes conjonctifs selon les relations entre des phrases : la coordination, la subordination, la subordination complétive et la relativation par suffixe conjonctif déterminatif.

²¹ La forme canonique <-eseo> qui est un suffixe conjonctif coordonnant de succession a plusieurs variantes selon des conditions phonologiques ou morphologiques :

(i) -eoseo /-aseo/-yeoseo/-seo
 (ii) -eo/-a/-yeo

verbes et que ces deux verbes sont considérés comme un verbe complexe composé, l'insertion du suffixe *-seo* qui signifie la succession est interdite. *-deul* (MPC) peut s'insérer entre deux verbes coordonnés de succession. Mais pour les autres types de coordination comme l'opposition, la causalité, etc., il faudra étudier les restrictions liées aux suffixes conjonctifs coordonnants²².

- (33) a. (*neohui-deul* + *E*), *bap* *meok-go-(deul* + *E*) *it-ni?*
 (vous + *E*) riz manger-**Sconj**-(**Mplu** + *E*) il y a-**Sti** ?
 (Est-ce que vous êtes en train de manger du riz ?)
- b. (*neohui-deul* + *E*), *geu chaek* *il-eo-(deul* + *E*) *bo-at-ni?*
 (vous + *E*) ce livre lire-**Sconj**-(**Mplu** + *E*) essayer-**Mps**-**Sti** ?
 (Est-ce que vous avez essayé de lire ce livre ?)

Le marqueur de pluriel *-deul* peut s'insérer entre un verbe et un verbe auxiliaire comme dans (33a) et (33b). Dans (33a), *-go* est un suffixe conjonctif reliant au verbe auxiliaire *itda* « il y a » et dans (33b), *-eo* au verbe auxiliaire *boda* « essayer ». *-go itda* « être en train de » et *-eo boda* « essayer de » marquent le sens aspectuel ou modal des verbes *meokda* « manger » et *ilda* « lire ».

Dans la grammaire coréenne, la notion d'auxiliaire est traditionnellement associée au niveau sémantique. Les éléments dits auxiliaires ajoutent un sens supplémentaire à la phrase (soit les verbes auxiliaires soit les postpositions auxiliaires). Les verbes auxiliaires ajoutent un sens aspectuel ou modal aux verbes principaux. Les suffixes conjonctifs *-eo*, *-ge*, *-ji* et *-go*²³ relient un verbe principal et un verbe auxiliaire : *-go sipda* « vouloir », *-go itda* « être en train de », *-eo boda* « essayer de », -

Les formes (i) sont des conjonctions de coordination. L'omission de *-seo* est possible ce qui donne (ii) : *dol-(aseo* + *a) gada* (= faire demi-tour et s'en aller). Les formes (ii) sans *-seo* peuvent s'employer comme conjonctions de subordination entre un verbe et un auxiliaire ou dans une combinaison figée des deux verbes : *dol-(a* + **aseo) gada* « mourir ». La forme *dol-a gada* est ambiguë sans contexte : « mourir » ou « faire demi-tour et s'en aller ».

²² Dans (a), *-go* est un suffixe conjonctif coordonnant de succession et dans (b) celui d'opposition :

- a. *bap meok-go-deul* *ga-geola*
 riz manger-**Sconj**-**Mplu** aller-**StImp**
 (Prenez du repas et allez)
- b. *geoli-e cha-deul-i* *o-go-(E + *deul)* *ga-nda*
 rue-**Postp.e** voiture-**Mplu**-**Nmtf** venir-**Sconj**-(*E* + **Mplu*) aller-**St**
 (Dans la rue, les voitures vont et viennent)

²³ Han S. H (2000) a considéré ces suffixes conjonctifs comme suffixes compléments parce qu'elles servent à introduire un complément phrastique.

ge hada « faire ; emploi causatif », *-ji malda* « arrêter de », etc. Il nous semble que *-deul* (MPC) peut s'insérer entre un verbe et un verbe auxiliaire sans restriction.

(ii) à l'intérieur d'un verbe ou d'un adjectif complexe²⁴

L'exemple de (34) appartient au cas des verbes composés :

- (34) (*neohui-deul* + *E*), *eo seo* *deul-eo-(deul* + *E)* *o-neola*
 (vous + *E*) vite entrer-**Sconj**-(**Mplu** + *E*) venir-**StImp**
 (Rentrez vite)

Le verbe *deul-eo-oda* « rentrer » dans (34) illustre un cas de verbe figé composé des deux verbes (*deulda* « entrer » et *oda* « venir ») et de *-eo*, un suffixe conjonctif subordonnant. Dans (34), l'insertion du suffixe *-seo* qui signifie la succession est interdite. La forme *deul-eo oda* n'est pas ambiguë et n'a que le sens figé : *deul-eo oda* « rentrer +*entrer et puis venir ». (35) appartient aux cas des adjectifs dérivés.

- (35) a. *salam-deul-i* *ganggyeong-(deul* + *E)* *ha-get-eo*
 homme-**Mplu-Nmtf** inflexible-(**Mplu** + *E*) être-**Mmod-St**
 (Il me semble que ces hommes sont inflexibles)
- b. *so-deul-i* *tteuntteun-(deul* + *E)* *hae-yo*
 vache-**Mplu-Nmtf** robuste-(**Mplu** + *E*) être-**St**
 (Les vaches sont robustes)

L'élément *ganggyeong* « inflexible » de (35a) et l'élément *tteuntteun* « robuste » de (35b) sont des radicaux *Rx* et forment des adjectifs par suffixation en *-hada*.

Regardons le cas où *-deul* (MPC) s'insère entre un nom prédicatif et le suffixe verbal *-hada*.

- (36) a. (*neohui-deul* + *E*), *gongbu-(deul* + *E)* *ha-geola*
 (vous + *E*) étude-(**Mplu** + *E*) faire-**StImp**
 (Faites des études)
- b. *geu ai-deul-eun* *gongbu-(deul-eul* + *E-leul)* *jal* *ha-nda*

²⁴ Le verbe (ou l'adjectif) complexe couvre le verbe (ou l'adjectif) dérivé et composé.

- ce enfant-**Mplu-Aux** étude-(**Mplu-Acc** + E-**Acc**) bien faire-**St**
 (Ces enfants font bien au moins leurs études)
- c. **geu ai-deul-eun gongbu-leul-deul jal ha-nda*
 *ce enfant-**Mplu-Aux** étude-**Acc-Mplu** bien faire-**St**

Dans (36a), *gongbu* « étude » est un nom prédicatif et il forme un verbe complexe²⁵ avec *-hada*. Etant donné que c'est un nom non comptable, *gongbu* « étude » ne peut pas se combiner avec *-deul*, lorsqu'il est sujet. Dans la combinaison avec d'autres postpositions nominales, il se place à droit du nom et à gauche de la postposition nominale ((36b) - (36c)). Il nous paraît que l'insertion de MPC à l'intérieur d'un verbe ou d'un adjectif complexe n'a pas de restrictions liées à valeur lexicale des verbes ou aux suffixes verbaux : *tago-nada* (doué - être = être doué), *ala-deutda* (savoir - entendre = comprendre), *bulhaeng-hada* (malheureux - être = être malheureux), *joa-hada* (amoureux - être = être amoureux), *nakseo-hada* (griffonnage - faire = griffonner), *mokyuk-hada* (bain - faire = prendre un bain), *yok-meokda* (injurer - manger = recevoir des injures), *beol-batda* (punition - recevoir = recevoir une punition), etc.

(iii) Après le suffixe complémenteur

Nous allons considérer la distribution *-deul* avec le suffixe complémenteur (*-go*).

- (37) a. [*geu-neun geuleon il-eul jal ha-nda*]-*go-deul ha-yo*
 [il-**Nmtf** ce travail-**Acc** bien faire-**St**]-**Scomp-Mplu** faire-**St**
 (On dit qu'il fait bien ce genre du travail)
- a'. * [*geu-neun geuleon il-eul jal-deul ha-nda*]-*go ha-yo*
 * [il-**Nmtf** ce travail-**Acc** bien-**Mplu** faire-**St**]-**Scomp** faire-**St**
- b. *Ina-ya, ai-deul-hante [ppali bumonim-kke jeonhwa-hae-la]-go hae-la*
 Ina-**Voc** enfant-**Mplu-Postp.e** [vite parents-**Postp.e** téléphone-faire-**StImp**]-**Scomp**
 faire-**StImp**
 (Ina, dis aux enfants de téléphoner vite aux leurs parants)
- = b'. *Ina-ya, ai-deul-hante [ppali-deul bumonim-kke jeonhwa-hae-la]-go hae-la*
 Ina-**Voc** enfant-**Mplu-Postp.e** [vite-**Mplu** parents-**Postp.e** téléphone-faire-**StImp**]-

²⁵ Han S.H (2000, pp. 22-23) a considéré les verbes en *-hada* dérivés de substantifs comme des verbes complexes en *N-hada*. La relation d'équivalence entre $N_0 \ W \ N\text{-Acc} \ hada$ et $N_0 \ W \ [N\text{-hada}]$ est régulière et systématique aux niveaux morphologique, sémantique et syntaxique.

Scomp faire-**StImp**

(Ina, dis aux enfants de téléphoner vite aux leurs parants)

b''. *Ina-ya, ai-deul-hante [ppali bumonim-kke jeonhwa-hae-la]-go-**deul** hae-la

*Ina-**Voc** enfant-**Postp.e** [vite parents-**Postp.e** téléphone-faire-**StImp**]-**Scomp-Mplu**
faire-**StImp**

Dans ces phrases (37a) et (37b), les derniers *hada* sont interprétés comme « dire ». Dans (37a), le sujet du verbe est omis et *-deul* marque le pluriel du sujet omis. Le MPC associé au suffixe complémenteur dans (37a) ne peut pas être copié à l'intérieur de la complétive. En revanche, dans (37b) le sujet du prédicat *hada* est un nom propre *Ina*. Le MPC peut s'insérer dans la complétive, mais il ne peut se placer après le suffixe complémenteur *-go*.

Enfin, remarquons le cas un peu discutable où le MPC se place après les suffixes terminaux.

(38) a. *geuleom*, (*yeleubun-deul* + *E*), *swi-seyo-deul*.

Alors, (vous +E) se reposer-**StImp-Mplu**

(Alors, reposez-vous)

b. ?*geu-salam-deul-i* *sal-a* *it-eulkka-deul* ?
? ce-homme-**Mplu-Nmtf** vivre-**Sconj** il y a-**Sti-Mplu** ?

(Ces hommes sont-ils vivants ?)

b'. *geu-salam-deul-i* *sal-a-deul* *it-eulkka* ?
ce-homme-**Mplu-Nmtf** vivre-**Sconj-Mplu** il y a-**Sti**?

(Ces hommes sont-ils vivants ?)

Dans (38a), *-deul* après les suffixes terminaux marque le pluriel du sujet de la phrase. Les suffixes terminaux du mode impératif (38a) sont beaucoup plus naturels avec *-deul* que ceux du mode déclaratif ou interrogatif. En fait, le jugement d'acceptabilité de (38b) est douteux²⁶. Il nous semble que la combinaison entre le suffixe terminal *-da* et *-deul* est peu naturelle, mais celle entre le suffixe terminal *-yo*, *-ge*, etc. et *-deul* est plus naturelle. Les structures où *-deul* marque le pluriel du sujet de la phrase demandent une étude approfondie selon le type de phrase ou de suffixe terminal.

²⁶ cf. Certaines auteurs (cf. Kim Y. H (1988), Seo J. S (1996), Moon S. Ch (1995), Lee H. G (1991), etc.) acceptent tels exemples, tandis que Jun J. Y (1999) ne les acceptent pas.

3-2-4. Relations de paraphrase entre phrases à *-deul* qui marque le pluriel du sujet de la phrase

Lorsque *-deul* marque le pluriel du sujet de la phrase, il est parfois redondant et il arrive que les Cas 2, 3 ou 4 de la section 3-2-3 se combinent. Pour donner un effet emphatique ou certaine nuance pragmatique aux destinataires, il arrive que le locuteur répète *-deul* dans une même phrase, mais cette répétition ne modifie pas la relation d'équivalence syntaxique des phrases. Nous allons esquisser plus systématiquement les propriétés de cette relation suivant que l'apparition explicite du sujet est obligatoire ou non²⁷. Ici, nous excluons la relation discutable d'équivalence des phrases à *-deul* après les suffixes terminaux.

Type 1. N_0 -deul ($N_1 + E$) ($W + E$) V où N_0 = : nom comptable et N_1 = : nom non comptable

Nous présentons plusieurs exemples de relations transformationnelles entre des phrases où *-deul* (MPC) marque le pluriel du sujet de la phrase et où le sujet apparaît explicitement. Considérons les variantes d'une phrase à structure N_0 -deul N_1 W V.

- (39) a. *chingu-deul-i* *don-i* *eop-eoyo*
 ami-Mplu-Nmtf argent-Nmtf (ne pas avoir)-St
 (Les amis n'ont pas d'argent)
- = b. *chingu-deul-i* *don-deul-i* *eop-eoyo*
 ami-Mplu-Nmtf argent-Mplu-Nmtf (ne pas avoir)-St
 (Les amis n'ont pas d'argent)

Dans (39b), N_1 nom non comptable copie la pluralité de N_0 comme suit (i) .

$$\begin{array}{c} \text{[copie -deul après } N_1\text{]} \\ \text{(i) } N_0\text{-deul } N_1 \text{ W V} = N_0\text{-deul } N_1\text{-deul W V} \end{array}$$

Une relation analogue est illustrée dans l'exemple (40) qui a la structure N_0 -deul W V :

²⁷ Nous ne présentons que quelques exemples de relation de paraphrase entre phrases en *-deul*.

- (40) a. *chingu-deul-i* *ppali* *meok-eot-eo*
ami-**Mplu-Nmtf** vite manger-**Mpas-St**
(Les amis ont mangé vite)
- = b. *chingu-deul-i* *ppali-deul* *meok-eot-eo*
ami-**Mplu-Nmtf** vite-**Mplu** manger-**Mpas-St**
(Les amis ont mangé vite)

Dans (40b), *Adv* copie la pluralité de N_0 comme suit (ii) :

$$\text{[copie -deul après Adv]} \\ \text{(ii) } N_0\text{-deul} \quad W \quad Adv \quad V \quad = \quad N_0\text{-deul} \quad W \quad Adv\text{-deul} \quad V$$

Dans (41), le radical *Rx* forme un adjectif par suffixation en *-hada* ($N_0\text{-deul} \quad W \quad Rx\text{-hada}$) et l'insertion de *-deul* entre le radical *Rx* et le suffixe *-hada* est possible :

- (41) a. *so-deul-i* *tteuntteun* *hae-yo*
vache-**Mplu-Nmtf** robuste faire-**St**
(Les vaches sont robustes)
- = b. *so-deul-i* *tteuntteun-deul* *hae-yo*
vache-**Mplu-Nmtf** robuste-**Mplu** faire-**St**
(Les vaches sont robustes)

La variante de la phrase à structure $N_0\text{-deul} \quad W \quad Rx\text{-hada}$ s'établit comme suit (iii) :

$$\text{[copie -deul après Rx]} \\ \text{(iii) } N_0\text{-deul} \quad W \quad Rx\text{-hada} = N_0\text{-deul} \quad W \quad Rx\text{-deul} \quad hada$$

Nous vérifions par plusieurs exemples que la répétition ou le déplacement de *-deul* dans une même phrase ne modifie le sens des phrases que d'une manière stylistique, lorsque N_0 est comptable et pluriel, et que *-deul* marque le pluriel du sujet (N_0) de la phrase en plusieurs positions.

Type 2. Omission du sujet pluriel

En coréen, dans les phrases interrogatives, propositionnelles et impératives, le sujet

est facultatif. L'omission du sujet peut même rendre la phrase plus naturelle. Cela produit de nouvelles variantes de phrases où *-deul* marque le pluriel du sujet omis.

- (42) a. *neohui-deul*, *ppali* *o-neola*
 vous vite venir-**StImp**
 (Venez vite)
- = b. *neohui-deul*, *ppali-deul* *o-neola*
 vous vite-**Mplu** venir-**StImp**
 (Venez vite)
- = c. *ppali-deul* *o-neola*
 vite-**Mplu** venir-**StImp**
 (Venez vite)

(42b) est transformé en (42c) par l'omission de *N₀-deul* comme suit (iv) :

$$\begin{aligned} & [(N_0\text{-}deul) \text{ z. }] \\ \text{(iv)} \quad & N_0\text{-}deul \quad W \quad V = W \quad V \end{aligned}$$

À partir de (42a), on obtient (42b) par (ii) et (42c) par (iv) :

$$\begin{aligned} \text{(ii)} \quad & N_0\text{-}deul \quad Adv \quad V = N_0\text{-}deul \quad Adv\text{-}deul \quad V \\ \text{(iv)} \quad & N_0\text{-}deul \quad Adv\text{-}deul \quad V = Adv\text{-}deul \quad V \end{aligned}$$

Dans (42c), l'omission du sujet et le non-occurrence de *-deul* rend la phrase ambiguë quant au nombre sans contexte précis :

- (43) *ppali* *o-neola*
 vite venir-**StImp**
 ((Viens + Venez) vite)

Ces variantes sont aussi observées avec un nom prédicatif non comptable *gongbu*. Par exemple :

- (44) a. *neohui-deul*, *gongbu* *ha-geola*
 vous étude-**Mplu** faire-**StImp**
 (Faites des études)

- = b. *neohui-deul, gongbu-deul ha-geola*
 vous étude-**Mplu** faire-**StImp**
 (Faites des études)
- = c. *gongbu-deul ha-geola*
 vous étude-**Mplu** faire-**StImp**
 (Faites des études)

Dans (44b), *Npred* non comptable copie la pluralité de *N₀* comme suit (v) :

$$\text{[copie -deul après } N_{pred}\text{]} \\ \text{(v) } N_0\text{-deul } N_{pred} \text{ } V = N_0\text{-deul } N_{pred}\text{-deul } V$$

Les transformations entre ces trois phrases ((44a) - (44c)) peuvent être décrites de la même façon :

$$\begin{aligned} \text{(v) } N_0\text{-deul } W \text{ } N_{pred} \text{ } V &= N_0\text{-deul } W \text{ } N_{pred}\text{-deul } V \\ \text{(iv) } N_0\text{-deul } W \text{ } N_{pred}\text{-deul } V &= W \text{ } N_{pred}\text{-deul } V \end{aligned}$$

Nous observons la relation de paraphrase dans la structure un peu plus compliquée (*N₀-deul N₁ W V-Sconj Vaux*) :

- (45) a. *neohui-deul, bap meok-go it-ni?*
 vous riz manger-**Sconj** il y a-**Sti** ?
 (Est-ce que vous êtes en train de manger du riz ?)
- = b. *neohui-deul, bap meok-go-deul itni?*
 vous riz manger-**Sconj-Mplu** il y a-**Sti** ?
 (Est-ce que vous êtes en train de manger du riz ?)

La variante (45b) de la phrase (45a) à structure *N₀-deul W V-Sconj Vaux* s'établit comme suit (vi) :

$$\text{[copie -deul après } Sconj\text{]} \\ \text{(vi) } N_0\text{-deul } W \text{ } V\text{-Sconj } Vaux = N_0\text{-deul } W \text{ } V\text{-Sconj-deul } Vaux$$

(45b) produit la phrase équivalente (45c) par (iv), et (45c) est transformé en (45d) par (i) et par (iv) :

- b. *neohui-deul*, *bap* *meok-go-deul* *itni?*
 vous riz manger-**Sconj-Mplu** il y a-**Sti** ?
 (Est-ce que vous êtes en train de manger du riz ?)
- = c. *bap* *meok-go-deul* *itni?*
 riz manger-**Sconj-Mplu** il y a-**Sti** ?
 (Est-ce que vous êtes en train de manger du riz ?)
- = d. (*neohui-deul*+E), *bap-deul* *meok-go* *it-ni?*
 (vous + E) riz-**Mplu** manger-**Sconj** il y a-**Sti** ?
 (Est-ce que vous êtes en train de manger du riz ?)

(45d) est encore transformé en (45e) par (vi) et (iv) :

- e. (*neohui-deul*+E), *bap-deul* *meok-go-deul* *itni?*
 (vous + E) riz-**Mplu** manger-**Sconj-Mplu** il y a-**Sti** ?
 (Est-ce que vous êtes en train de manger du riz ?)

Avec le sujet facultatif, (45a) à structure $N_0\text{-deul } N_1 \text{ } W \text{ } V \text{ } Vaux$ a 7 phrases équivalentes. Nous excluons (46) qui n'a aucune occurrence de *-deul* dans les relations transformationnelles, parce que nous ne pouvons pas savoir s'il a le sens pluriel du sujet.

- (46) *bap* *meok-go* *itni?*
 riz manger-**Sconj** il y a-**Sti** ?
 (Est-ce que (tu es + vous êtes) en train de manger du riz ?)

Nous récapitulons les relations de paraphrase entre des phrases à $[N_0\text{-deul } N_1 \text{ } W \text{ } V]$:

$$\begin{aligned}
 N_0\text{-deul } N_1 \text{ } W \text{ } V &= N_0\text{-deul } N_1\text{-deul } W \text{ } V \\
 N_0\text{-deul } W \text{ } Adv \text{ } V &= N_0\text{-deul } W \text{ } Adv\text{-deul } V \\
 N_0\text{-deul } W \text{ } Rx\text{-hada} &= N_0\text{-deul } W \text{ } Rx\text{-deul } hada \\
 N_0\text{-deul } W \text{ } Npred \text{ } V &= N_0\text{-deul } W \text{ } Npred\text{-deul } V \\
 N_0\text{-deul } W \text{ } V\text{-Sconj } Vaux &= N_0\text{-deul } W \text{ } V\text{-Sconj-deul } Vaux \\
 N_0\text{-deul } W \text{ } V &= W \text{ } V
 \end{aligned}$$

Ces transformations ne sont pas orientées et nous pourrions y ajouter d'autres transformations après des études approfondies sur les contraintes de transformations.

La distribution de *-deul* est comparable à celle des postpositions auxiliaires²⁸. Les postpositions auxiliaires peuvent se combiner avec plusieurs éléments comme des noms, des adverbes, des postpositions nominales dites adverbiales, des suffixes complémentateurs, et s'insérer entre deux verbes ou à l'intérieur d'un verbe complexe. Elles ne donnent aucune fonction syntaxique aux éléments combinés et n'ajoutent qu'une nuance à la phrase. Nous considérons donc le morphème du pluriel *-deul* comme une postposition auxiliaire²⁹.

3-3. Combinatoire des postpositions nominales

3-3-1. Forme canonique, variantes phonologiquement conditionnées, variantes libres et homonymes

Comme nous l'avons mentionné dans le chapitre 3 (p. 72), certains éléments morphologiques peuvent avoir plusieurs formes en fonction de conditions phonologiques ou morphologiques. Par exemple, le morphème *-neun* qui correspond à l'expression *c'est...que* en français possède deux formes dont l'emploi obéit à des conditions phonologiques sur le contexte gauche :

-neun / V_____

-eun / C_____

On peut sélectionner une des deux formes comme forme canonique. Le choix de la forme canonique est arbitraire, mais il a généralement une motivation liée à l'organisation générale de la description. Comme des variations similaires sont observées avec d'autres suffixes, on a adopté une règle simple : la variante observée après voyelle est choisie comme forme canonique. Ainsi, *-neun*, *-ya*, *-leul*, *-wa*, *-lo*, etc.

²⁸ D'autres postpositions auxiliaires se combinent avec les radicaux *Rx*.

a. *so-deul-i* *tteuntteun-do* *hae-yo*
vache-**Mplu-Nmtf** robuste-**Aux** faire-**St**
(Les vaches sont aussi robustes)

b. *i* *jip-i* *joyong-eun* *ha-da*
ce maison-**Nmtf** calme-**Aux** être-**St**

(Cette maison est au moins calme, mais on ne peut rien dire d'autre sur elle)

²⁹ Dans certains dictionnaires comme *Yonsei Hangukeo Sajeon* (*Yonsei* Dictionnaire du Coréen), on étiquette *-deul* comme un suffixe pluriel dans tous ses emplois. Dans le dictionnaire *Uli mal keun Sajeon* (Grand Dictionnaire du Coréen), on a traité *-deul* comme une postposition nominale, et dans le dictionnaire *Pyojun Gugkeo Daesajeon* (Grand Dictionnaire Standard du Coréen) (1999), comme deux entrées : un suffixe pluriel et une postposition auxiliaire.

sont des formes canoniques. Chacune de ces formes est en distribution complémentaire avec une ou plusieurs formes morphologiquement apparentées.

D'autre part, il existe aussi des variantes interchangeables sans modification de sens. Nous avons considéré les variantes comme libres si les trois conditions suivantes sont réunies :

- (i) Les deux morphèmes appartiennent à la même classe de postpositions
- (ii) La combinatoire avec les autres postpositions est la même
- (iii) Le sens ne change pas en cas de substitution, ou change seulement d'une manière stylistique

Par exemple :

- (47) a. *na-neun chingu-ege-neun geojinmal-eul anha-nda*
 Je-**Nmtf** ami-**Postp.e-Aux** mensonge-**Acc** ne pas faire-**St**
 (C'est à mon ami que je ne mentis pas)
- b. *na-neun chingu-ege-n geojinmal-eul anha-nda*
 Je-**Nmtf** ami-**Postp.e-Aux** mensonge-**Acc** ne pas faire-**St**
 (C'est à mon ami que je ne mentis pas)
- c. *na-neun chingu-ege-man-eun geojinmal-eul anha-nda*
 Je-**Nmtf** ami-**Postp.e-Aux-Aux** mensonge-**Acc** ne pas faire-**St**
 (C'est seulement à mon ami que je ne mentis pas)

La relation entre *-neun* (47a) et *-eun* (47c) est phonologiquement conditionnée. En revanche, la relation entre *-neun* (47a) et *-n* (47b) est libre : *-neun* et *-n* sont interchangeables sans modification de sens. On utilise souvent *-n* dans la langue orale.

Il existe des postpositions homonymiques ; elles ont la même forme. Etant donné que les propriétés, notamment les combinaisons avec d'autres postpositions, sont différentes, on doit les traiter séparément. Par exemple, *-eseo* peut être une postposition dite adverbiale ou une postposition du nominatif, s'il apparaît avec des noms humains collectifs.

- (48) a. *jeongbu-eseo usuha-n haksang-deul-ege janghakgeum-eul ju-nda*
 gouvernement-**Nmtf** être excellent-**Sfd** étudiant-**Mplu-Postp.e** bourse-**Acc** donner-**St**
 (Le gouvernement donne les bourses aux étudiants excellents)
- b. *Minu-neun doseogwan-eseo gongbu-leul ha-nda*

Minu-**Nmtf** bibliothèque-**Postp.eseo** étude-**Acc** faire-**St**
 (Minu étudie dans la bibliothèque)

-*kkaji* peut être une postposition casuelle dite adverbiale ou une postposition auxiliaire. Les graphes de combinaisons des postpositions auxiliaires et des postpositions adverbiales sont différents. Comparons (49a) et (49a') avec (49b) et (49b') :

- (49) a. *ai-deul-kkaji* *geu* *sasil-eul* *a-nda*
 enfant-**Mplu-Aux** ce fait-**Acc** savoir-**St**
 (Même les enfants savent ce fait)
- a'. **ai-deul-kkaji-neun* *geu* *sasil-eul* *a-nda*
 *enfant-**Mplu-Aux-Aux** ce fait-**Acc** savoir-**St**
- b. *beoseu unjeonja-deul-ui* *paeop-ttaemune* *hakgyo-kkaji* *geoleoga-t-da*
 chauffeur d'autobus-**Mplu-Gén** à cause de grève école-**Postp.kkaji** aller à pied-**Mpas-St**
 (Je suis allé **jusqu'à** l'école à pied à cause de la grève des chauffeurs d'autobus)
- b'. *beoseu unjeonja-deul-ui* *paeop-ttaemune* *hakgyo-kkaji-neun* *geoleoga-t-da*
 chauffeur d'autobus-**Mplu-Gén** à cause de grève école-**Postp.kkaji-Aux** aller à pied-**Mpas-St**
 (Je suis allé **au moins jusqu'à** l'école à pied à cause de la grève des chauffeurs d'autobus)

3-3-2. Combinatoire des postpositions nominales

Une fois que l'on dispose de la liste des variantes de toutes les formes canoniques, il faut décrire systématiquement toutes les combinaisons de postpositions nominales. Ces combinaisons peuvent être représentées dans un transducteur fini. Or, lors de la construction du dictionnaire des séquences de postpositions nominales sous forme de transducteur fini, nous devons considérer à la fois deux sortes de contraintes combinatoires : celles entre postpositions nominales et celles entre la postposition nominale et le nom. Quand il s'agit des contraintes combinatoires entre postpositions nominales, il suffit de considérer les conditions phonologiques et les variantes libres. En revanche, pour les contraintes combinatoires entre la postposition nominale et le nom, il s'agit non seulement de conditions phonologiques mais aussi de conditions sur les traits sémantiques.

Rappelons notre classification des séquences nominales figées selon les

contraintes combinatoires avec les postpositions nominales (chapitre 3, p. 72). Il y a neuf classes des séquences nominales figées : N11, N12, N13, N21, N22, N23, N31, N32 et N33.

- N11 : Terminaison avec consonne sauf *l*, (+Hum, +Anm, +CollHum)
(3400 séquences de postpositions nominales)
- N12 : Terminaison avec consonne sauf *l*, (+InstHum)(1650)
- N13 : Terminaison avec consonne sauf *l*, (+Plt, +Inanim)(1900)
- N21 : Terminaison avec consonne *l*, (+Hum, +Anm, +CollHum)(3400)
- N22 : Terminaison avec consonne *l*, (+InstHum)(1650)
- N23 : Terminaison avec consonne *l*, (+Plt, +Inanim)(1900)
- N31 : Terminaison avec voyelle, (+Hum, +Anm, +CollHum)(3400)
- N32 : Terminaison avec voyelle, (+InstHum)(1650)
- N33 : Terminaison avec voyelle, (+Plt, +Inanim)(1900)

Il existe six traits sémantiques (+Hum, +Anm, +CollHum, +Plt, +Inanim, +InstHum) et trois contextes gauches en fonction de la condition phonologique :

- (i) après consonne sauf *l* (C\L____)
- (ii) après consonne *l* (L____)
- (ii) après voyelle (V____)

Pour construire les graphes (N11.grf ou N33.grf), tout d'abord, nous avons groupé toutes les séquences de postpositions nominales d'après les conditions phonologiques et nous les avons regroupées en 8 classes d'après les traits sémantiques :

Classe	Conditions phonologiques	Traits sémantiques
P1	C\L____, L____	+Hum, +Anm, +CollHum, +Plt, +Inanim, +InstHum
P2	V____	+Hum, +Anm, +CollHum, +Plt, +Inanim, +InstHum
P3	C\L____, L____, V____	+Hum, +Anm, +CollHum, +Plt, +Inanim, +InstHum
P4	V____, L____	+Hum, +Anm, +CollHum, +Plt, +Inanim, +InstHum
P5	C\L____	+Hum, +Anm, +CollHum, +Plt, +Inanim, +InstHum
P6	C\L____, L____, V____	+Hum, +Anm, +CollHum
P7	C\L____, L____, V____	+Plt, +Inanim
P8	C\L____, L____, V____	+InstHum

Figure 8. Classes de postpositions nominales

Par exemple, les séquences de postpositions nominales dans la classe P5 satisfont à une condition : elles se combinent avec les séquences nominales figées qui se terminent par une consonne autre que *l*. Les classes P1, P2, P4 et P5 sont phonologiquement exclusives, et P6, P7 et P8 sémantiquement exclusives. Les séquences de postpositions nominales de la classe P3 se combinent avec tous les noms.

En fait, nous avons fortement réduit le nombre des séquences de postpositions nominales combinables avec les séquences nominales figées à l'aide des traits sémantiques ci-dessus (+Hum, +InstHum, +Inanim, etc.). Nous estimons qu'une classification plus raffinée des séquences nominales figées peut faire encore réduire le nombre des séquences de postpositions nominales combinables avec les noms. Par l'étude de la combinatoire syntactico-sémantique entre noms et postpositions nominales, il faudrait améliorer la classification des séquences nominales figées.

Nous présentons un exemple de graphe de postpositions nominales « N11.grf », qui est compatible avec les séquences nominales figées de la classe N11 (figure 9). Dans la figure 9, le morphème pluriel *-deul* étiqueté *Pl* se place après les séquences nominales et avant les autres postpositions nominales. Puisqu'il est terminé par la consonne *l*, nous faisons apparaître la classe P4 après *-deul* et la classe P5 dans la branche sans *-deul*. Les formes de P4 et de P5 sont différentes selon les conditions phonologiques. Les noms comptables et non comptables peuvent se combiner avec *-deul*, même si le rôle de *-deul* soit différent selon les distributions. Lorsqu'il est placé entre les séquences nominales et les postpositions nominales, avec un nom comptable, *-deul* marque le pluriel du substantif, et avec un nom non comptable, le pluriel du sujet de la phrase. Pour l'interprétation précise de la forme avec *-deul* ou sans *-deul*, il faudrait obligatoirement distinguer les noms comptables des noms non comptables. *-deul* peut aussi apparaître après certaines postpositions nominales dites adverbiales pour marquer le pluriel du sujet de la phrase (figure 10, p. 140)³⁰.

³⁰ Comme nous l'avons examiné dans la section 3-2, la distribution de *-deul* et l'interprétation des formes sans *-deul* ou avec *-deul* sont très compliquées. De plus, nous ne pouvons pas répondre immédiatement au jugement d'acceptabilité sur les séquences comme *ai-deul-hago-deul* (enfant-Mplu-Postp.hago-Mplu)(cf. note 8, p.115). Il nous semble que nous sommes dans la même situation hésitante pour les séquences contenant les noms non comptables : *gongbu-deul-lo-deul* (étude-Mplu-Postp.lo-Mplu). Si l'allocutaire est pluriel, on pourrait accepter de telles formes dans certains contextes. Pour le moment, nous avons compté de telles formes comme possibles dans nos graphes de postpositions nominales. En outre, nous avons traité de la même façon la combinatoire entre les noms comptables accompagnés de *-deul* et les autres postpositions nominales, et entre les noms non comptables

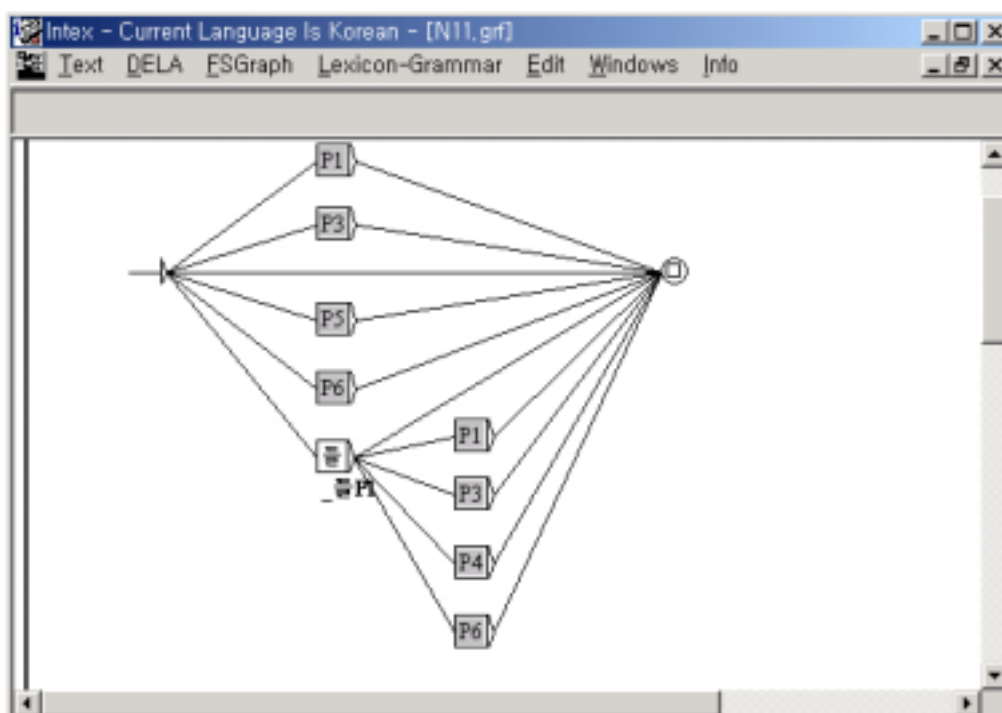


Figure 9. N11.grf

Regardons le sous-graphe « P5.grf » dans lequel nous trouvons les postpositions nominales et leurs étiquettes avec la représentation à l'aide de transducteurs (figure 10, p. 140). Ici, il faut faire une remarque technique. En fait, le codage de la flexion nominale dans INTEX est moins lisible que dans la figure 10. Notre texte utilise le graphe de la figure 10 pour faciliter l'exposition. Citons un exemple de DELACF dans INTEX pour le codage sur la flexion du français :

aime,aimer.V:P1s:P3s:S1s:Y2s

Le caractère « : » est un caractère spécial pour INTEX. Pour indiquer les homonymies, on utilise « : ». Après « : », on décrit les traits flexionnels de chaque entrée. Par exemple, « P » signifie le présent et « 1 » indique première personne et « s » singulier, etc. Chaque valeur de trait flexionnel est obligatoirement représenté par un caractère unique.

En coréen, nous avons considéré les étiquettes des postpositions nominales (formes canonique et catégorie) comme des traits flexionnels. Cela nous a obligé à

accompagnés de *-deul* et les autres postpositions nominales. Nous laissons pour une prochaine étude la description plus élaborée de la combinatoire entre les noms et *-deul*, et entre les noms non comptables accompagnés de *-deul* et les autres positions nominales.

coder toutes les étiquettes de postpositions nominales par caractères ASCII. Par exemple, dans *jip-eseo-neun.jip.N:jG* (maison-**Postp.eseo**-**Aux**) la postposition **-eseo** (**Postp.eseo**) est marquée par « j » et la postposition auxiliaire **-neun** (**Aux**) par « G ».

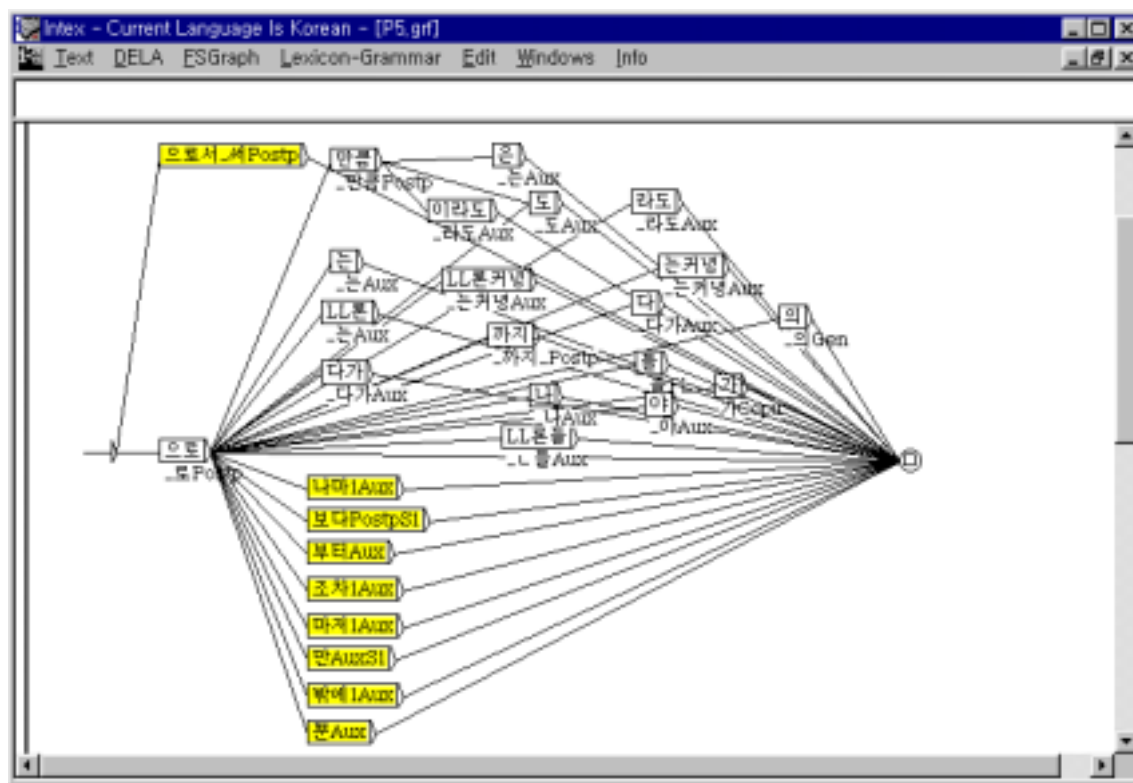


Figure 10. Le sous-graphe P5. grf de N11

Voici une glose qui associe à chaque code ASCII la forme canonique et la classe linguistique d'une postposition nominale.

A<ga, Nmtf>	B<ga, Copu>	C<nama, Aux>	D<na, Aux>
E<na, Conj>	F<neunkeunyeng, Aux>	G<neun, Aux>	H<ni, Conj>
I<deunga, Conj>	J<deunji, Conj>	K<deun, Conj>	L<lago, Aux>
M<lado, Aux>	N<mattana, Postp>	O<laymyeon, Aux>	P<laya, Aux>
Q<lan, Aux>	R<lang, Postp>	S<lang, Conj>	T<lya, Conj>
U<leul, Acc>	V<myeo, Conj>	W<myeon, Conj>	X<ya, Aux>
Y<wa, Conj>	Z<wa, Postp>	a<lo, Postp>	b<loseo, Postp>
c<losseo, Postp>	d<ya, Voc>	e<yeo, Voc>	f<ndeul, Aux>
g<njeuk, Aux>	h<llnang, Aux>	i<da, Aux>	j<eseo, Postp>
k<eseo, Nmtf>	l<daga, Aux>	m<kkeseo, Nmtf>	n<egeseo, Postp>
o<go, Conj>	p<ege, Postp>	q<e, Postp>	r<hago, Postp>
s<kke, Postp>	t<deoleo, Postp>	u<bogo, Postp>	v<gachi, Postp>

w<kkajigo, <i>Postp</i> >	x<kkajina, <i>Postp</i> >	y<kkaji, <i>Aux</i> >	z<kkaji, <i>Postp</i> >
0<deul, <i>Pl</i> >	1<ui, <i>Gen</i> >	2<hamyeo, <i>Conj</i> >	3<jocha, <i>Aux</i> >
4<do, <i>Aux</i> >	5<ttala, <i>Aux</i> >	6<manyang, <i>Postp</i> >	7<mada, <i>Aux</i> >
8<cheoleom, <i>Postp</i> >	9<majeo, <i>Aux</i> >	_<manchi, <i>Postp</i> >	=<mankeum, <i>Postp</i> >
@<man, <i>Aux</i> >	#<malgo, <i>Aux</i> >	\$<bakke, <i>Aux</i> >	^<bodam, <i>Postp</i> >
&<buteo, <i>Aux</i> >	*<buteo, <i>Postp</i> >	(<ppun, <i>Aux</i> >)<boda, <i>Postp</i> >
~<hago, <i>Conj</i> >	%<daelo, <i>Postp</i> >	[<chigo, <i>Aux</i> >]<keonyeng, <i>Aux</i> >
<tolok, <i>Postp</i> >	?<kkaena, <i>Aux</i> >	<yamalo, <i>Aux</i> >	'<ladeunga, <i>Conj</i> >
/<ladeunji, <i>Conj</i> >			

Avec ces conventions, la figure 10 est codée dans INTEX comme suit :

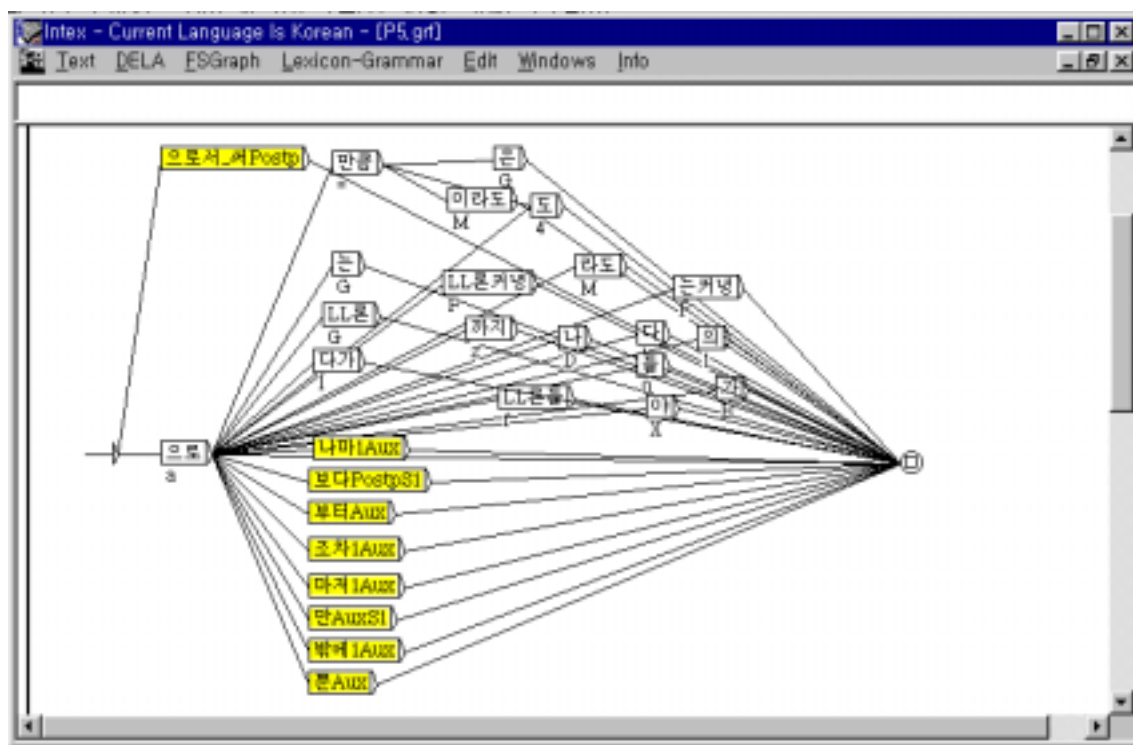


Figure 11. Le sous-graphe P5. grf de N11 dans INTEX en coréen

Malgré cette différence de codage, les informations sur les postpositions nominales sont les mêmes que dans la figure 10.

Dans la description par des graphes, nous utilisons l'opérateur d'effacement « L » (*Left*)³¹ d'INTEX. Avant de l'expliquer, introduisons le fait que l'écriture coréenne

³¹ Pour les plus détails sur les trois opérateurs (« L » *Left*, « R » *Right* et « C » *Copy*) dans INTEX, voir la section 1-1. Génération automatique des formes fléchies des séquences nominales figées du chapitre 5.

est fondée sur une notion graphique de syllabe. En coréen, la syllabe est l'unité de base de l'écriture et une syllabe écrite est constituée de l'une des cinq manières suivantes:

- (i) une consonne initiale (parmi 19) + une voyelle (21) + une consonne finale (28)
- (ii) une consonne initiale (19) + une voyelle (21)
- (iii) la consonne muette (1) + une voyelle (21)
- (iv) une consonne (28)
- (v) une voyelle (21)

Dans (i) - (v), le chiffre entre parenthèses indique le nombre de lettres pratiquement employées dans les textes coréens. Les trois cas (i) - (iii) sont les formes générales de la structure syllabique du coréen. Le cas (iii) est un cas où il n'y a pas de consonne prononcée en tête de la syllabe. Une syllabe constituée d'une seule consonne ou d'une seule voyelle comme dans (iv) et (v) est utilisée dans des cas très spéciaux pour citer une consonne (ou une voyelle). Exemple :

- (50) *i* *bueok-eun* ㄴ-*ja-hyeong-i-da*
 ce cuisine-**Aux** ㄴ-lettre-forme-être-**St**
 (Cette cuisine est en L)

La figure 12 montre comment nous traitons une modification graphique qui se produit lors de la concaténation de deux postpositions nominales. La forme *lon* est obtenue en concaténant les deux postpositions nominales *-lo* et *-n*. La syllabe graphique constituée d'une consonne *n* devient la consonne finale de la syllabe *lon* « ㄴ », puisque l'on n'écrit pas *lo n* « ㄴ로 » en deux syllabes dans les textes coréens. Ce phénomène peut facilement être décrit par l'opérateur d'effacement « L ».



Figure 12. Exemple de modification graphique dans l'écriture syllabique

Un L efface un octet. En coréen, toute syllabe de l'un des types (i) - (v) prend deux octets et l'effacement d'une syllabe en coréen demande donc deux L. Deux L effacent *lo* et *lon* le remplace.

4. Conclusion

Pour profiter des avantages de la représentation par graphes, nous avons proposé la représentation sous la forme de graphes pour certains types de séquences nominales et pour les postpositions nominales. Spécialement, pour la description des séquences de postpositions nominales par des graphes dans INTEX, nous avons étudié leur classification linguistique et leur contraintes combinatoires. En outre, nous avons examiné les deux emplois du morphème de pluriel *-deul*. Les conditions d'emploi du morphème *-deul* sont complexes. Lorsqu'il se combine avec un substantif comptable, il marque le pluriel du substantif lui-même. Si ce morphème se combine avec un substantif non comptable ou avec diverses catégories grammaticales comme un adverbe, une postposition nominale dite adverbiale, un suffixe conjonctif, un suffixe complémentateur, etc., il marque le pluriel du sujet de la phrase.

Le graphe explicite la relation d'équivalence qui lie les formes. Il permet donc de choisir un représentant unique, de présenter ses variantes et ses étiquettes linguistiques à l'aide de transducteurs et de décrire les modifications graphiques lors de la concaténation d'éléments.

Chapitre 5. Applications du dictionnaire électronique des séquences nominales figées et de leurs formes fléchies

Dans ce chapitre, nous présentons une méthode de génération des formes fléchies des séquences nominales figées, puis nous appliquons notre dictionnaire et nous examinons ses avantages dans plusieurs domaines du traitement automatique des langues naturelles.

Dans la section 1, à partir de codes flexionnels comme N11 ou N33 du dictionnaire électronique et de la fonctionnalité de flexion du menu DELA d'INTEX (« Inflection »), nous engendrons toutes les formes fléchies des séquences nominales figées. Nous fusionnons deux bases de données : les séquences nominales figées et les séquences de postpositions nominales. Dans la section 2, nous justifions la validité de notre méthode en appliquant notre dictionnaire des formes fléchies des noms figés à espacement facultatif (NFF) à des textes coréens. Dans la section 3, nous examinons l'application de notre dictionnaire électronique dans plusieurs domaines du traitement automatique des langues naturelles comme la recherche d'informations, l'extraction de mots-clés, l'analyse syntaxique et la traduction automatique.

1. Construction automatique du dictionnaire des formes fléchies des séquences nominales figées avec INTEX

1-1. Génération automatique des formes fléchies des séquences nominales figées

Avant d'aborder l'étape de la génération automatique des formes fléchies des séquences nominales figées du coréen, nous présentons brièvement la méthode de production des formes fléchies des mots simples en français : il s'agit de la méthode mise au point au LADL par B. Courtois (1990) et M. Silberztein (1993). DELAS est une liste de tous les mots simples sous leurs formes lemmatisées : par exemple, l'infinitif pour les verbes, le masculin singulier pour les adjectifs, etc. Chacun des mots simples est accompagné d'un code flexionnel qui permet d'engendrer automatiquement toutes les formes fléchies à partir de sa forme lemmatisée. Par exemple, pour l'entrée du DELAS français (M. Silberztein, 1996 et A. Chrobot, 2001) :

nouveau,A72

L'exemple ci-dessus représente le lemme *nouveau* qui est un adjectif. Voici le

transducteur de flexion A72 représenté par le graphe au format INTEx. Il est associé aux adjectifs du type *nouveau* :

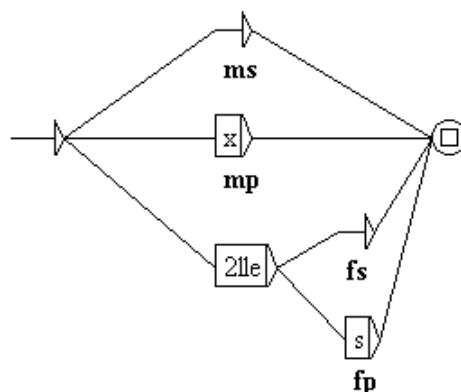


Figure 1. Transducteur de flexion A72

Puisque le masculin singulier est équivalent au lemme, le suffixe vide est ajouté au lemme. Mais pour produire le pluriel au masculin, il faut ajouter le suffixe « x ». Pour produire le féminin singulier, il faut effacer les deux dernières lettres du lemme et ajouter la terminaison « lle ».

Introduisons les opérateurs de pile dans INTEx. L'effacement du dernier caractère est représenté par l'opérateur « L » (*Left*) et on peut abréger une séquence d'opérateur d'effacement en indiquant simplement le nombre d'effacements : « 2 » est équivalent à « LL ». Si on efface les deux dernières lettres du lemme et qu'on ajoute « lles », on produit la forme du pluriel au féminin. Le transducteur de flexion A72 dans le dossier « Inflection » génère les quatre entées suivantes du DELAF :

```
nouveau,nouveau.A72:ms
nouvelle,nouveau.A72:fs
nouveaux,nouveau.A72:mp
nouvelles,nouveau.A72:fp
```

Les quatre formes fléchies sont associées au même lemme *nouveau* après une virgule. Les propriétés linguistiques suivent le point. Les deux points « : » introduisent les traits morphologiques des formes fléchies obtenues. INTEx contient deux autres opérateurs « R » (*Right*) et « C » (*Copy*) pour décrire plus facilement le changement systématique d'une voyelle dans un ensemble de mots de base qui ne partagent pas la même désinence : « R » permet de sauter une lettre vers la droite dans le lemme, et « C »

permet de la dupliquer. Par exemple, la flexion de la première ou de la troisième personne du singulier présent (:*Pls:P3s*) des verbes comme *acheter*, *geler* ou *mener* est représentée grâce à la commande suivante :

LLLLRèCe (ou de façon abrégée : ***4RèCe***)

En se plaçant à la fin du mot, il faut reculer de 4 lettres à gauche, aller d'une lettre à droite *R*, insérer la lettre *è*, recopier *C* la lettre courante *n*, et insérer la lettre *e* :

Opération	Lemme	Pile	Résultat
	mener^		mener
4	m^ener	ener	m
R	me^ner	ner	m
è	me^ner	ner	mè
C	men^er	er	mèn
e	men^er	er	mène

Ici, « L » correspond à l'opérateur de pile « empiler » (PUSH), « R » à l'opérateur de pile « dépiler » (POP) et « C » à l'opérateur de pile « dépiler et imprimer ». La flexion d'un dictionnaire DELAS est immédiate : chaque opérateur prend un temps constant ; la construction de chaque forme prend un temps proportionnel à longueur du suffixe du transducteur ; la flexion de chaque entrée prend un temps proportionnel au nombre de chemins du transducteur (M. Silberztein, 2000)¹.

Maintenant, revenons au problème de génération des formes fléchies des

¹ Pour le programme sur le processus de la commande, voir Silberztein (1996, p.85) :

```

char *inflect (char *lemma, char *commands) {
static int ic, ir, il ; static char result [512] ;
strcpy (result, lemma);
for (ic=0, ir=il=strlen (res); commands[ic];ic++) {
switch (commands[ic]) {
case 'C': result [ir++]=lemma [il++];break;
case 'L': il--; ir--;break;
case 'R': il++;break;
default: result[ir++]=commands[ic];
}
}
result[ir]='\0';
return result;
}

```

séquences nominales figées du coréen dans INTEX. Dans le chapitre 3, nous avons examiné le codage de chaque entrée selon les contraintes combinatoires avec les postpositions nominales. Dans le chapitre 4, nous avons étudié la construction du dictionnaire des séquences de postpositions nominales sous forme de graphes. Pour construire un dictionnaire électronique exploitable, il existe une étape indispensable : la fusion de deux bases de données. En effet, des formes fléchies des séquences nominales figées apparaissent dans les textes coréens. A partir de codes flexionnels comme N11 ou N33 d'un dictionnaire électronique et de la fonctionnalité de flexion du menu DELA d'INTEX (« Inflection »), nous engendrons toutes les formes fléchies des séquences nominales figées. Par exemple :

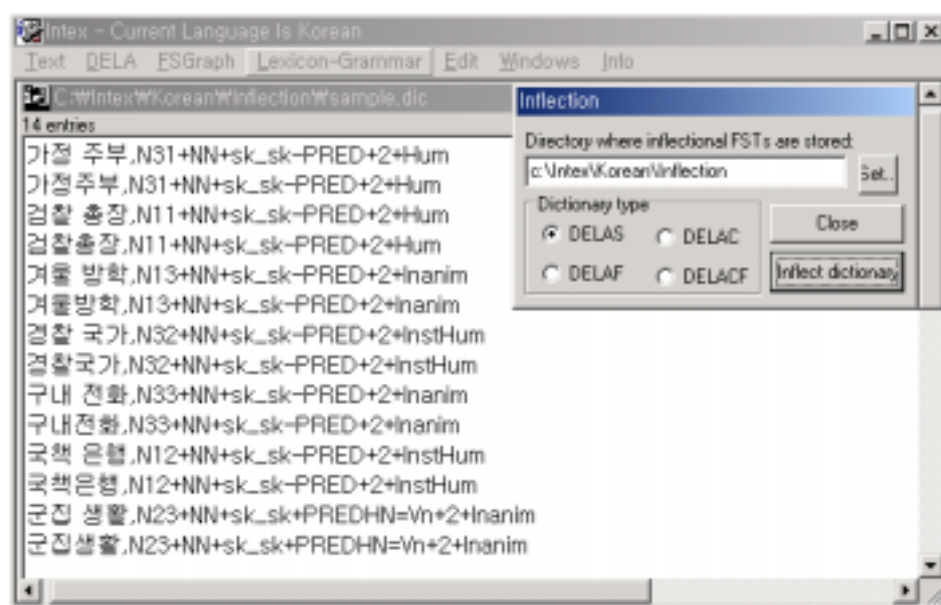


Figure 2. Dictionnaire du type DELAS pour la flexion

Cette fonctionnalité permet la flexion automatique en appelant des transducteurs de postpositions nominales. Le nom du transducteur est exactement le code associé à chaque entrée. Lorsqu'il s'agit de la flexion des séquences nominales du coréen, le module de flexion est suffisant pour engendrer tous les types de séquences nominales figées : les postpositions nominales ou les séquences de postpositions nominales se combinent toujours avec la dernière syllabe des séquences nominales figées, sans qu'aucun changement de forme n'intervienne au niveau de cette dernière syllabe.

Voici par exemple une entrée de nom figé à espacement facultatif et ses formes fléchies. Chaque exemple ci-dessous est comparable aux entrées du DELAC (1a) et du

DELACF ((1b) et (1c)) dans le système DELA (M. Silberztein, 1990).

(1) a. *gukchaek ^ eunhaeng*, N12+NN+sk_sk -PRED+2+InstHum

(politique-national - banque = banque nationale; Banque de France, Banque de Corée, etc.)

Dans (1a), le code « ^ » représente l'espacement facultatif et le code flexionnel N12 appelle le graphe de séquences de postpositions nominales « N12.grf ». À partir de « *gukchaek eunhaeng*, N12 », les formes fléchies ((1b) et (1c)) de (1a) sont automatiquement engendrées à l'aide du module de flexion d'INTEX.

b. *gukchaek**eunhaeng****deulmaneun***, *gukchaek eunhaeng*.

N12+NN+sk_skPRED+2+InstHum: *_deul***PI***_man***Aux***_neun***Aux**²

c. *gukchaek eunhaeng****deulmaneun***, *gukchaek eunhaeng*.

N12+NN+sk_sk-PRED+2+InstHum: *_deul***PI***_man***Aux***_neun***Aux**

Dans (1b) et (1c), les formes *gukchaek**eunhaeng****deulmaneun*** et *gukchaek eunhaeng****deulmaneun*** sont associées au même lemme *gukchaek eunhaeng* après une virgule. Les propriétés linguistiques suivent le point. Les deux points « : » introduisent les informations sur les postpositions nominales. Chaque postposition nominale est délimitée par « _ ». Les formes *deul*, *man*, *neun* en italique sont les formes canoniques des postpositions nominales. Chaque forme en gras **PI** (morphème pluriel) et **Aux** (Auxiliaire) représente une classe linguistique de postpositions nominales³.

1-2. Résultats

Avec les graphes des postpositions nominales, nous engendrons toutes les formes fléchies des séquences nominales figées. La figure 3 montre un échantillon très réduit des listes produites avec INTEX. La génération des formes fléchies des séquences nominales figées avec INTEX respecte en particulier l'usage coréen en ce qui concerne :

² Le codage après les deux points n'est pas tout à fait celui utilisé dans INTEX, mais les informations sont les mêmes. Pour faciliter l'exposition, nous marquons les formes canoniques en caractères coréens et les classes des postpositions nominales sous la forme des étiquettes *Postp*, *Aux*, etc. En fait, avec le logiciel INTEX chaque forme canonique accompagnée de l'étiquette d'une classe de postpositions nominales est représentée par un caractère ASCII. Pour ce codage, voir la note 19 dans le chapitre 4.

³ Pour la classification linguistique, voir la section 3-1 du chapitre 4.

- les variantes phonologiques et leurs formes canoniques
- les variantes libres et leurs formes canoniques
- la contraction graphique des formes non syllabiques (cf. figure 12 dans le chapitre 4)

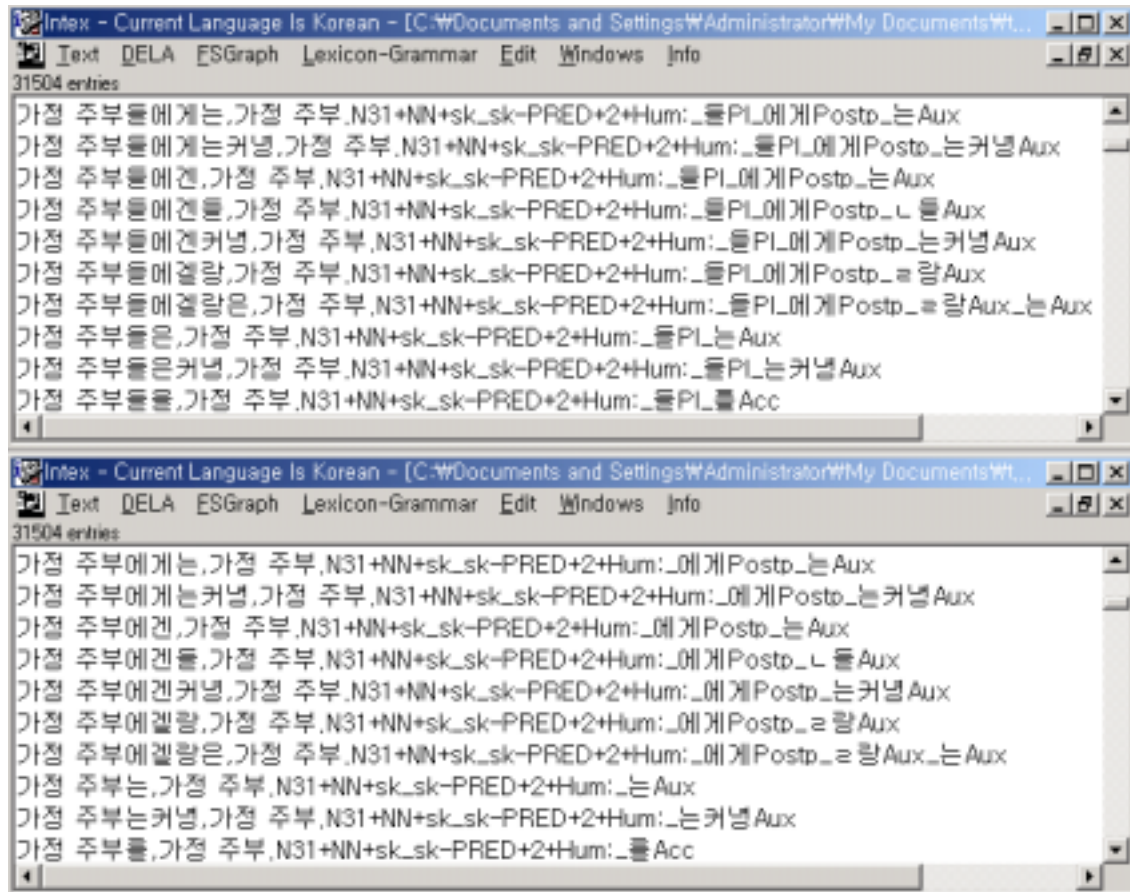


Figure 3. Résultat de la génération des formes fléchies des séquences nominales figées

Les quelques exemples ci-dessous illustrent ces trois difficultés.

- gajeong jubu-deul-eun* : *_deulPI_neunAux*
- gajeong jubu-neun* : *_neunAux*
- gajeong jubu-deul-ege-neun* : *_deulPI_egePostp_neunAux*
- gajeong jubu-deul-ege-n* : *_deulPI_egePostp_neunAux*

Les variations phonologiques sont visibles si l'on se réfère aux formes canoniques. Par exemple, dans (2a) et (2b) nous voyons les deux variantes phonologiques de la

postposition nominale <-neun> : -eun après la consonne *l* et -neun après voyelle. Quant aux variantes libres, qui apparaissent dans un contexte phonologique identique (par exemple, -neun et -n se trouvent après une voyelle et sont des formes d'une même postposition auxiliaire <-neun>), nous leur faisons correspondre la même forme canonique, **_egePostp_neunAux** pour les séquences de postpositions nominales -ege-neun et -ege-n ((2c) et (2d)). Nous observons aussi une contraction graphique : la forme *egen* est une forme graphiquement modifiée à partir de -ege et de la forme non syllabique -n. La syllabe graphique constituée d'une consonne *n* devient la consonne finale de la syllabe précédente *ge* (cf. figure 12 dans le chapitre 4).

Nous pouvons faire une estimation du nombre de formes fléchies de certaines classes, domaines techniques non compris :

Séquences nominales figées		Séquences de <i>Postp</i>	Séquences nominales figées avec des <i>Postp</i>
NC Simples	15000	2300	35×10^6
Autres NC	?	2300	?
NFF	45000 x 3	2300	311×10^6
NFO	?	2300	?

Figure 4. Estimation du nombre des formes fléchies des séquences nominales figées

Nam Jee-Sun (1994) a recensé 15000 noms compacts simples. Nous avons recensé 45000 noms figés à espacement facultatif. Ce nombre d'entrées doit être multiplié par 3, le nombre moyen de variantes liées à l'espacement. Par exemple, pour un nom figé à espacement facultatif constitué de deux noms (*NN*) le nombre de variantes est de deux. Pour trois noms (*NNN*) il est de quatre, pour quatre noms (*NNNN*) de huit, etc. Il existe des contraintes combinatoires entre séquences nominales figées et postpositions nominales. Le nombre moyen de séquences de postpositions pour un nom est environ 2300. Le nombre des séquences des postpositions pour un nom dépend du nom (cf. la section 2-2 du chapitre 3 et la section 3-3-2 du chapitre 4).

Calculons plus précisément le nombre des formes fléchies des NFF de

notre dictionnaire :

		Nombre de NFF		Nombre des variantes selon l'espacement	Nombre de formes fléchies
NN	N11, N21, N31	2087	42487	2087 x 2	2087 x 2 x 3400 = 14191600
	N12, N22, N32	1239		1239 x 2	1239 x 2 x 1650 = 4088700
	N13, N23, N33	39161		39161 x 2	39161 x 2 x 1900 = 148811800
NNN	N11, N21, N31	148	3277	148 x 4	148 x 4 x 3400 = 2012800
	N12, N22, N32	342		342 x 4	342 x 4 x 1650 = 2257200
	N13, N23, N33	2787		2787 x 4	2787 x 4 x 1900 = 21181200
NNNN	N11, N21, N31	4	202	4 x 8	4 x 8 x 3400 = 108800
	N12, N22, N32	20		20 x 8	20 x 8 x 1650 = 264000
	N13, N23, N33	178		178 x 8	178 x 8 x 1900 = 2705600
NNNNN	N32	2	7	2 x 16	2 x 16 x 1650 = 52800
	N13, N33	5		5 x 16	5 x 16 x 1900 = 152000
NNNNNN	N13	2	2	2 x 32	2 x 32 x 1900 = 121600
Total		45975		99674	195948100 ($\cong 200 \times 10^6$)

Figure 5. Nombre de NFF et de leurs formes fléchies⁴

Le nombre d'entrées doit être multiplié par le nombre d'espaces et de séquences de postposition nominales : par exemple, 2087 (entrées) x 2 (variantes liées à l'espacement) x 3400 (séquences de postpositions nominales) indique le nombre des formes fléchies de NFF constitué de deux noms (NN) appartenant aux classes N11, N21 et N31.

1-3. Compression du DELAF et du DELACF de NFF en FST déterministes minimaux

Une fois construit sous la forme d'un DELAF d'INTEX, le dictionnaire de toutes les formes fléchies des NFF, il est placé dans le dossier DELACF de la langue coréenne avec l'extension « .dic ». Or, nous obtenons deux types de formes fléchies de NFF en coréen : les formes sans aucun espace et celles avec au moins un espace. Nous pouvons

⁴ Par exemple, les NFF des classes N11, N21 ou N31 prennent le même nombre de séquences de postpositions nominales.

tout mettre dans le dossier DELACF, mais nous avons mis les formes fléchies des NFF sans espace dans le dossier DELAF et celles des NFF avec espace dans le dossier DELACF pour faciliter l'exposition (cf. nous allons comparer le nombre de NFF reconnus par DELAF et celui par DELACF (figure 15, p. 162).

Mais, comme nous l'avons vu dans la figure 4, le nombre des formes fléchies des NFF est 200×10^6 et la taille du fichier qui est environ 12 Go est trop énorme pour que le dictionnaire soit utilisé directement dans INTEX. Il est avantageux de convertir ce dictionnaire en transducteur, dont la taille typique est de l'ordre de quelques Mo⁵. INTEX construit automatiquement le transducteur équivalent au dictionnaire DELAF ou DELACF à l'aide du module « Compress into FST » du menu **DELA**. Le processus de compression dans INTEX consiste en deux phases⁶ :

- (i) La liste textuelle d'entrées du DELAF est convertie en un arbre lexicographique (ou un cas particulier d'automate déterministe acyclique⁷), où l'état initial est la racine de l'arbre et les états terminaux sont des feuilles. La figure 7 présente un arbre lexicographique construit pour un dictionnaire contenant six mots anglais : *ant*, *ants*, *apse*, *apses*, *pat* et *pats*. Le langage reconnu⁸ par l'automate de la figure 7 est : $L(A)^9 = \{ant, ants, apse, apses, pat, pats\}$.

⁵ Par exemple, le DELAF français qui contient environ 700 000 formes a une taille de plus de 30 Mo et le DELAF russe qui contient 3,5 millions de formes a une taille de plus de 100 Mo (M. Silberztein, 2000, p.161).

⁶ (...) *compact* enters each line from *stdin* into a deterministic finite state automaton, and then minimize it. The minimization process is in O (number of states), because the automaton has no cycle (M. Silberztein 1996, p. 86).

⁷ Un *automate fini* A est défini par la donnée d'un quintuplet d'ensembles (B, Q, I, T, F) (D.Revuz, 1991) :

- un alphabet fini B ;
- un ensemble fini d'états noté Q ;
- une partie I de Q d'états initiaux ;
- une partie T de Q d'états terminaux ;
- un ensemble fini F de triplets (p, a, q) , où p et q sont des états et a un symbole de l'alphabet B ; ces triplets sont les transitions de l'automate.

Les *automates déterministes* sont des automates dans lesquels pour tout état p de Q et pour tout symbole a de B , il existe au plus une transition partant de l'état p étiquetée par a . On demande en plus que I soit réduit à un unique élément, c'est-à-dire que l'automate ne possède qu'un seul état initial. On parle d'*automates acycliques* lorsqu'il n'existe pas de suite de transitions dans l'automate qui permette de passer deux fois par le même état.

⁸ Un mot est reconnu par un automate si et seulement s'il est l'étiquette d'un chemin qui part de l'état initial et arrive à un état terminal. On lit le mot symbole après symbole, tout en se déplaçant dans l'automate en suivant les transitions étiquetées par le symbole courant. Si après avoir lu tout le mot on se trouve dans un état terminal de l'automate, le mot est reconnu ; sinon le mot n'est pas reconnu (M. Silberztein, 1989).

⁹ On définit le langage reconnu par un automate, noté $L(A)$, comme l'ensemble des mots reconnus.

pats}

- (ii) L'arbre lexicographique est minimisé¹⁰ selon la méthode basée sur la propriété d'équivalence (E) d'états dans un automate. Deux états p et q sont équivalents ($(p, q) \in E$), si et seulement si leurs langages droits sont égaux, le langage droit d'un état étant l'ensemble de tous les suffixes reconnus à partir de cet état jusqu'à l'un des états finals. Par exemple, dans la figure 7 les ensembles d'états qui ont les mêmes langages droits sont $\{5, 9, 13\}$, $\{4, 8, 12\}$, et $\{3, 11\}$. Si tous les états équivalents sont fusionnés, nous obtenons l'automate minimal comme la figure 8.

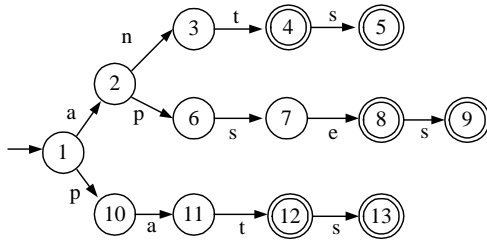


Figure 6. Un arbre lexicographique¹¹

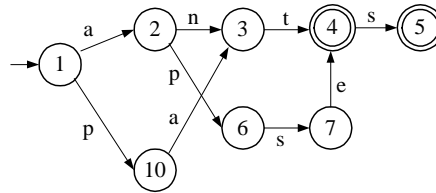


Figure 7. Un automate minimal¹²

L'utilisation d'un arbre lexicographique permet d'obtenir des vitesses intéressantes, puisque la vitesse est dépendante de la hauteur, et non plus directement du nombre d'entrées. Mais leur mise en œuvre nécessite un espace important. Ce point pose un problème pour la compression des formes fléchies pour les langues agglutinantes. Nous y reviendrons plus loin (pp. 156-157).

Dans un DELAF compressé sous INTEX, aucun symbole de sortie n'est attaché aux transitions. La production de l'étiquette grammaticale ne se fait qu'après avoir atteint l'état final. La minimisation est effectuée par l'algorithme ci-dessus. La seule différence est que les états terminaux sont distingués par les différentes productions qui leur sont attribuées. L'état final est associé au numéro qui représente l'indexe de l'information lexicale correspondante dans un tableau. Pour réduire le nombre des informations lexicales, les lemmes sont remplacés par les commandes utilisés lors du processus de la flexion dans INTEX. Si un lemme est égal à son entrée, il est effacé. Voici par exemple trois entrées du DELAF (M. Silberztein 1996, p. 74) :

¹⁰ L'ensemble des automates reconnaissant un langage donné admet un plus petit élément par le nombre d'états ; cet automate le plus petit est appelé *automate minimal*, qui est unique. Des algorithmes de minimisation transforment un automate donné en l'automate minimal qui reconnaît le même langage.

¹¹ A. Chrobot 2001, p. 38.

¹² A. Chrobot 2001, p. 38.

aider,aider.V:W
aidions,aider.V:I1p:S1p
volions,voler.V:I1p:S1p

Elles sont représentées comme suit :

aider,.V:W
aidions,4er.V:I1p:S1p
volions,4er.V:I1p:S1p

Par conséquent, les deux lignes suivantes sont stockées dans le fichier d'informations :

.V:W
4er.V:I1p:S1p

Le transducteur dans INTEx est donc représenté par deux fichiers : un fichier « .bin » qui contient les transitions et un fichier « .inf » qui contient les informations lexicales que l'on trouve dans le dictionnaire. Extrayons un dictionnaire français représenté par un transducteur (M. Silberztein, 2000, p.162) :

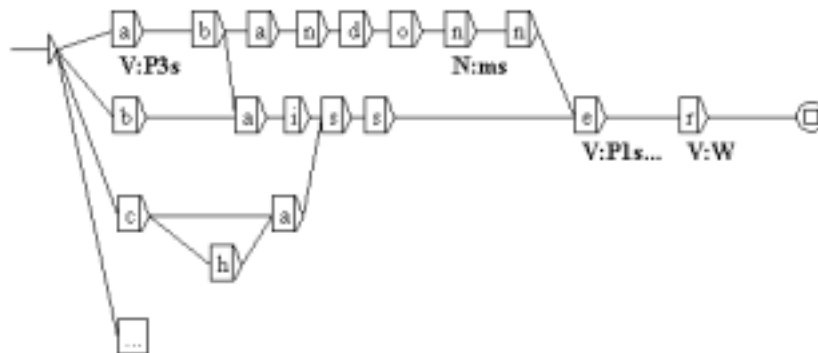


Figure 8. Extrait d'un dictionnaire représenté par un transducteur

Tous les nœuds qui contiennent une production sont traités comme des nœuds terminaux. Ce transducteur reconnaît la forme *a*, le nom *abandon*, les verbes à l'infinitif *abandonner*, *abaisser*, *casser* et *chasser*, les formes conjuguées *abandonne*, *abaisse*, *casse* et *chasse* appartenant au dictionnaire textuel d'origine. Ce transducteur est déterministe minimale, parce que les préfixes et les suffixes de toutes les formes y sont mis en facteur. Plus le dictionnaire contient des entrées avec les préfixes et les suffixes

communs, plus le taux de compression d'un dictionnaire textuel vers un automate est élevé. Le temps d'accès est linéaire en fonction de la longueur du mot recherché.

Or, comme nous l'avons remarqué ci-dessus, le processus de compression en deux phases dans INTEx exige un espace mémoire suffisant au moins pour contenir l'arbre lexicographique qui résulte de la première phase (i) (p. 152). Ceci pose un problème de performance pour de gros dictionnaires, et surtout pour des langues à flexion très riche, car l'arbre lexicographique peut s'avérer plus grand que la taille de mémoire disponible. Pour améliorer ce problème, des méthodes plus efficaces sont élaborées : la méthode de D.Revuz (1991) et celle de J.Daciuk et al. (2000).

D.Revuz (1991) a proposé un algorithme qui permet, à partir d'un dictionnaire textuel d'obtenir un automate *pseudo-minimal* de taille très réduite par rapport à un arbre lexicographique correspondant (p. 152). Cet algorithme procède à une première minimisation, en comprimant certains suffixes communs à plusieurs mots. Les mots sont ajoutés progressivement dans l'ordre lexicographique inverse : la comparaison de lettre se fait de droite à gauche. Pour chaque nouveau mot, il faut trouver le plus long suffixe commun de ce mot avec tous les mots qui sont déjà dans l'automate. Ce suffixe commun est factorisé. L'insertion du préfixe du nouveau mot, qui reste après l'isolation du suffixe commun, se fait de gauche à droite en utilisant les états déjà existants pour un seul prédécesseur ou en les dédoublant pour plus d'un prédécesseur (D. Revuz 1991, pp. 42-45). Après avoir obtenu l'automate pseudo-minimal, la phase de minimisation a lieu : pour tous les éléments de la partition par hauteur croissante, trier les états par leurs transitions et s'ils sont équivalents, les fusionner. La minimisation est basée sur la propriété d'équivalence de langages droits (cf. voir la phase (ii), p. 153). La complexité de l'algorithme est linéaire en fonction de taille du lexique pour la première phase, et linéaire en fonction de taille du nombre d'états dans l'automate pseudo-minimal pour la deuxième phase.

J. Daciuk et al. (2000) a présenté une méthode plus efficace. Cette méthode ne contient qu'une seule phase. Les mots sont rajoutés dans l'ordre lexicographique de telle façon qu'après chaque nouvelle insertion l'automate obtenu reste minimal. Pour un mot à insérer, il faut d'abord trouver le plus long préfixe commun avec les mots qui sont déjà dans l'automate. A partir du dernier état de ce préfixe, il faut parcourir le chemin du mot précédent (de droite à gauche), et vérifier si certains de ses états ne peuvent pas être remplacés par des états équivalents (du point de vu de leurs langages droits) qui existent déjà dans l'automate. Finalement, le suffixe du mot courant est ajouté à l'automate, et la procédure recommence pour un mot suivant. Reprenons un exemple de construction incrémentale d'un automate minimal pour le même ensemble de mots que ceux de la

figure 6 (A. Chrobot, 2001, p. 39-40) :

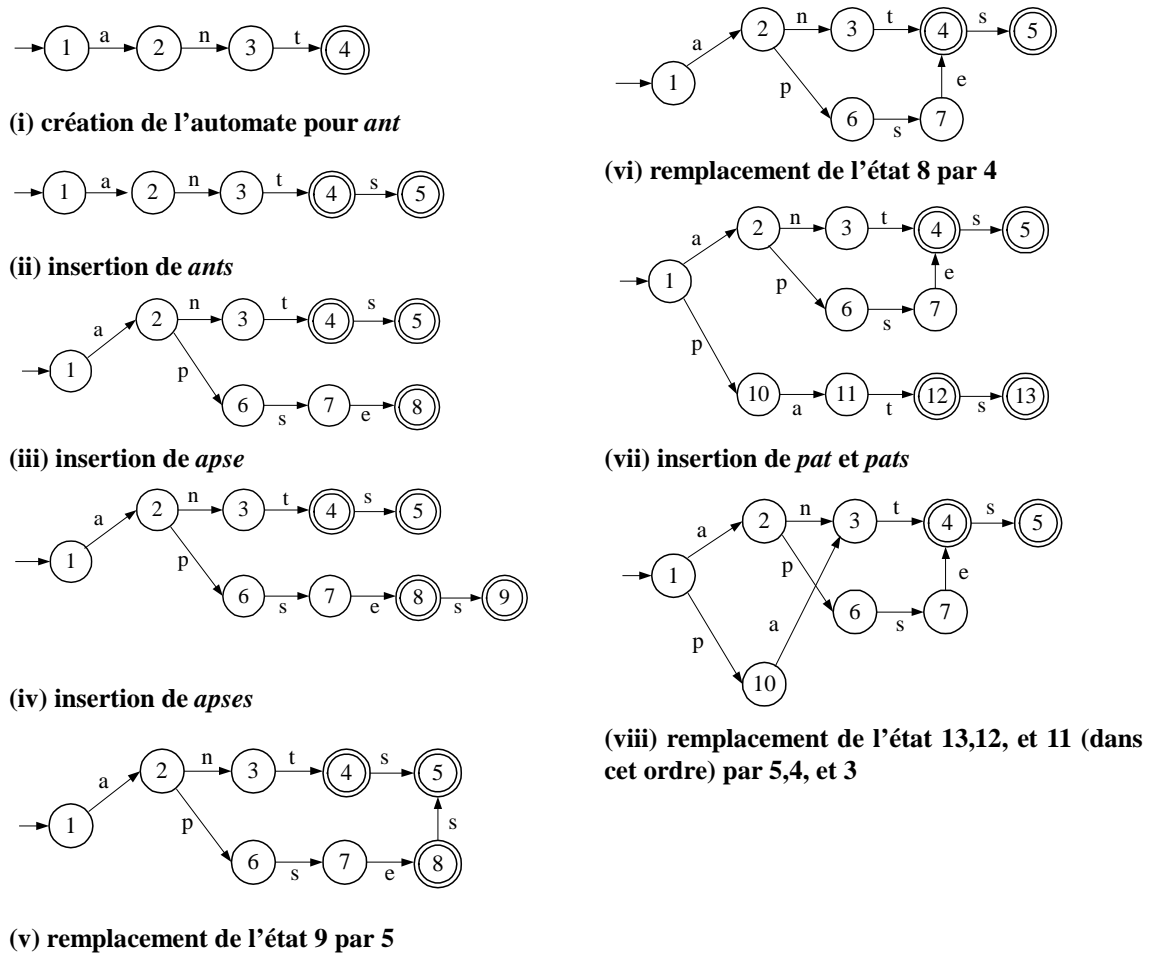


Figure 9. Construction incrémentale d'un automate minimal

Maintenant, revenons au problème de la compression des formes fléchies des séquences nominales figées du coréen avec INTEX. L'algorithme d'INTEX pour la compression en deux étapes (pp. 152-153) n'est pas convenable pour celle du DELAF du coréen à la taille trop importante en raison du problème du mémoire, car la fabrication de l'automate déterministe acyclique à partir d'une liste textuelle demande une quantité de mémoire suffisante pour le contenir. Selon notre expérience, alors qu'INTEX peut engendrer les formes fléchies textuelles à quelques Go, il ne peut pas compresser un dictionnaire des formes fléchies textuelles d'une taille supérieure environ 200 Mo en une seule fois. En fait, puisque les algorithmes d'INTEX ne sont pas conçus au départ pour des langues agglutinantes comme le coréen, il existe des défauts de la procédure lors du traitement du coréen avec INTEX. Mais nous les avons utilisés, parce

que notre objectif était seulement de tester les données et les méthodes linguistiques que nous avons élaborées dans notre thèse. En utilisant INTEX pour le coréen, nous sommes allée jusqu'à la limite de la flexibilité permise par INTEX en attendant la réalisation d'une version coréenne d'INTEX. En prenant la méthode de Daciuk et al. (2000), on pourrait améliorer le problème de mémoire lors de la compression des très gros dictionnaires. Cependant, il faut noter que toutes les méthodes que nous avons envisagées partent d'une liste de formes fléchies. Pour la compression du coréen, mieux vaudrait probablement construire le dictionnaire de formes fléchies comprimé directement à partir du dictionnaire de formes canoniques, sans passer par la liste de formes fléchies qui est trop grosse.

Puisqu'INTEX ne peut pas compresser un dictionnaire des formes fléchies textuelles d'une taille supérieure environ 200 Mo en une seule fois, nos dictionnaires DELAF (5.3 Go) et DELACF (6.5 Go) ont une taille trop importante pour être compressés en une seule fois. Nous avons donc décidé de subdiviser, avant génération, le fichier des séquences nominales figées en plusieurs fichiers : 31 fichiers de NFF sans espace et 38 fichiers de NFF pour les noms figés avec espace. Chaque fichier d'environ 170 Mo est réduit à 4.3 Mo (fichier « .bin ») et à 12 Mo (fichier « .inf ») par la compression. Par conséquent, nous avons mis le dictionnaire comprimé des formes fléchies des NFF sans espace dans le dossier DELAF et celui avec au moins un espace dans le dossier DELACF¹³ :

- 31 fichiers x 4.3 Mo de « .bin » = 133 Mo et
31 fichiers x 12 Mo de « .inf » = 372 Mo dans le dossier DELAF
- 38 fichiers x 4.3 Mo de « .bin » = 163 Mo
38 fichiers x 12 Mo de « .inf » = 456 Mo dans le dossier DELACF

Nous récapitulons notre méthode de construction des dictionnaires des formes fléchies des NFF avec INTEX :

¹³ Plus exactement,
 31 fichiers x 4.45 Mo de « .bin » \cong 138 Mo et
 31 fichiers x 11.97 Mo de « .inf » \cong 371 Mo dans le dossier DELAF
 38 fichiers x 4.13 Mo de « .bin » \cong 157 Mo
 38 fichiers x 12.1 Mo de « .inf » \cong 460 Mo dans le dossier DELACF

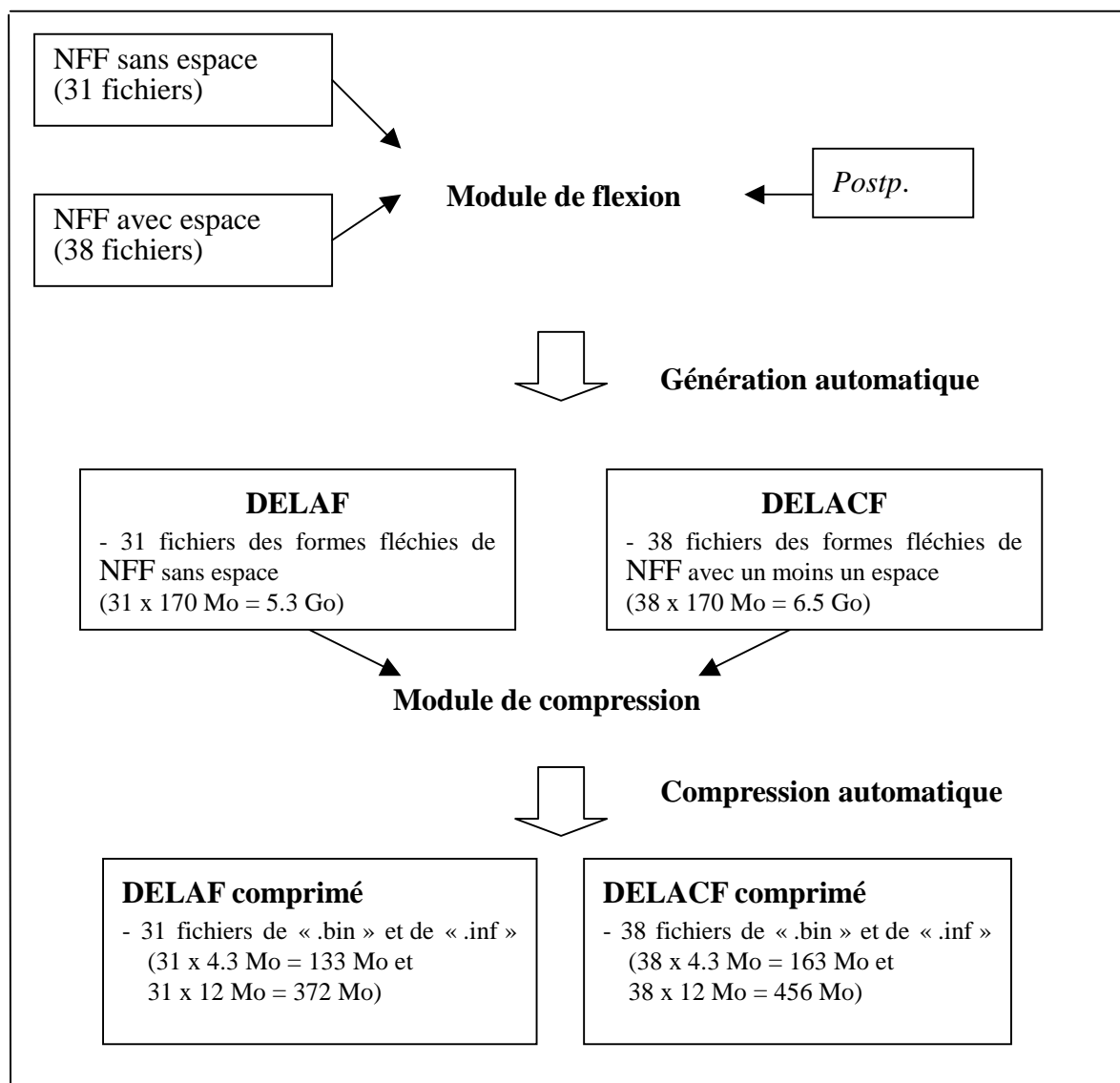


Figure 10. Construction automatique du dictionnaire des formes fléchies de NFF à l'aide d'INTEX

2. Reconnaissance automatique des NFF par dictionnaire électronique

2-1. Reconnaissance des NFF avec INTEX

Pour la reconnaissance des NFF dans les textes coréens par la consultation de notre dictionnaire, nous utilisons le système INTEX et le CD-ROM de corpus écrit du Projet Sejong 21 (<http://www.sejong.or.kr>).

Or, lors de la consultation du dictionnaire avec INTEX, nous ne pouvons pas

lancer les 69 fichiers « .bin » en même temps ni par l'interface graphique ni par la commande en ligne parce que le nombre de fichiers « .bin » dans le dossier DELACF est limité à 10 fichiers par consultation. Pour le moment, lorsque nous utilisons notre dictionnaire avec INTEX, nous appliquons les quatre parties de DELAF et de DELACF avec le même texte :

- 1^{ière} partie : « 1sflx.bin ~ 10sflx.bin » et « 1cflx.bin ~ 10cflx.bin »
- 2^{ième} partie : « 11sflx.bin ~ 20sflx.bin » et « 11cflx.bin ~ 20cflx.bin »
- 3^{ième} partie : « 21sflx.bin ~ 30sflx.bin » et « 21cflx.bin ~ 30cflx.bin »
- 4^{ième} partie : « 31sflx.bin » et « 31cflx.bin ~ 38cflx.bin »

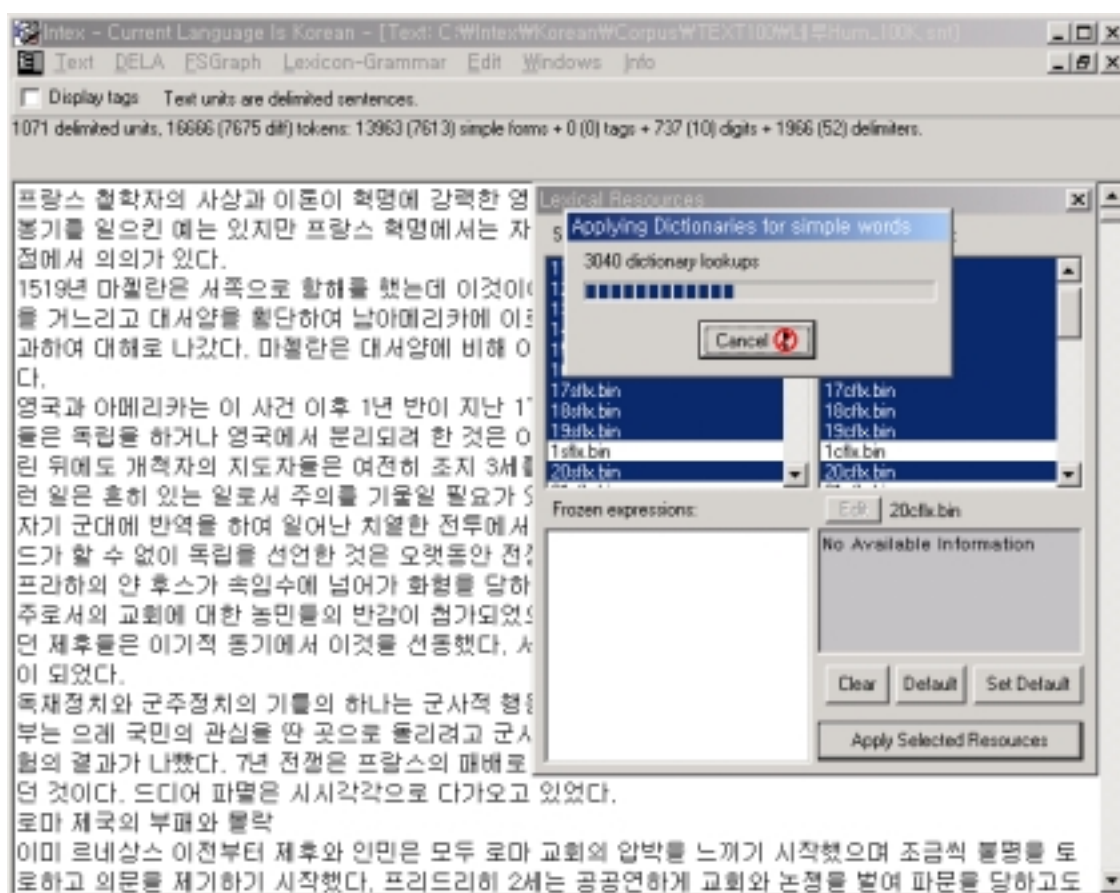


Figure 11. Application de la 2^{ième} partie de DELAF et de DELACF

Nous présentons un échantillon réduit des listes reconnues par la consultation de la 2^{ième} partie du DELAF et du DELACF. Le texte utilisé a une taille de 100 Ko et il s'agit d'un livre sur l'histoire du monde.

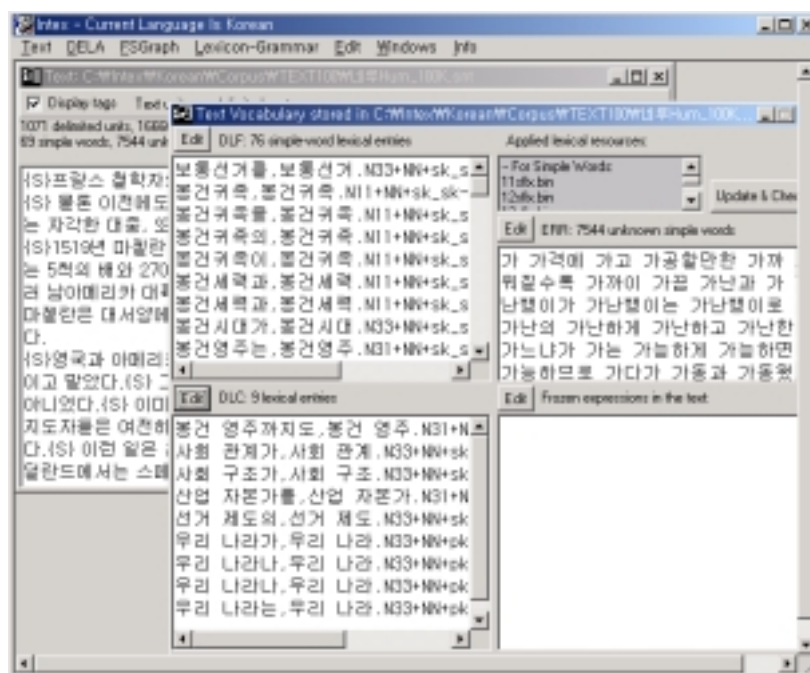


Figure 12. Résultat de consultation de la 2^{ème} partie de DELAF et de DELACF

INTEX peut construire concordance des NFF reconnus par la consultation du dictionnaire à l'aide du module « Locate Pattern ». Lors de cette opération, nous utilisons la variable <N> qui fait référence aux noms décrits dans notre dictionnaire.

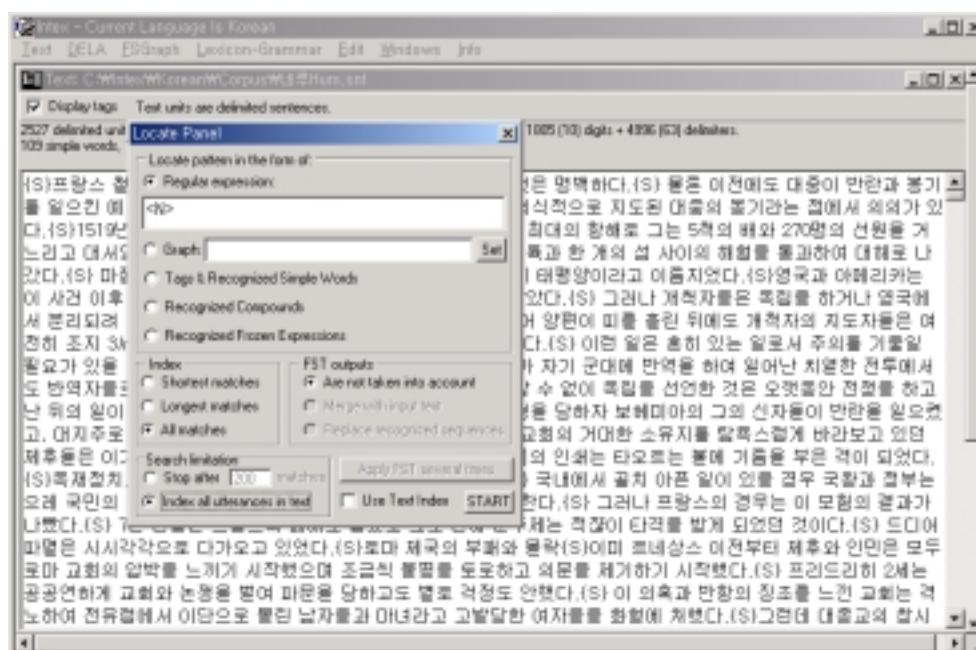


Figure 13. Recherche de NFF dans le texte à l'aide de la variable <N>

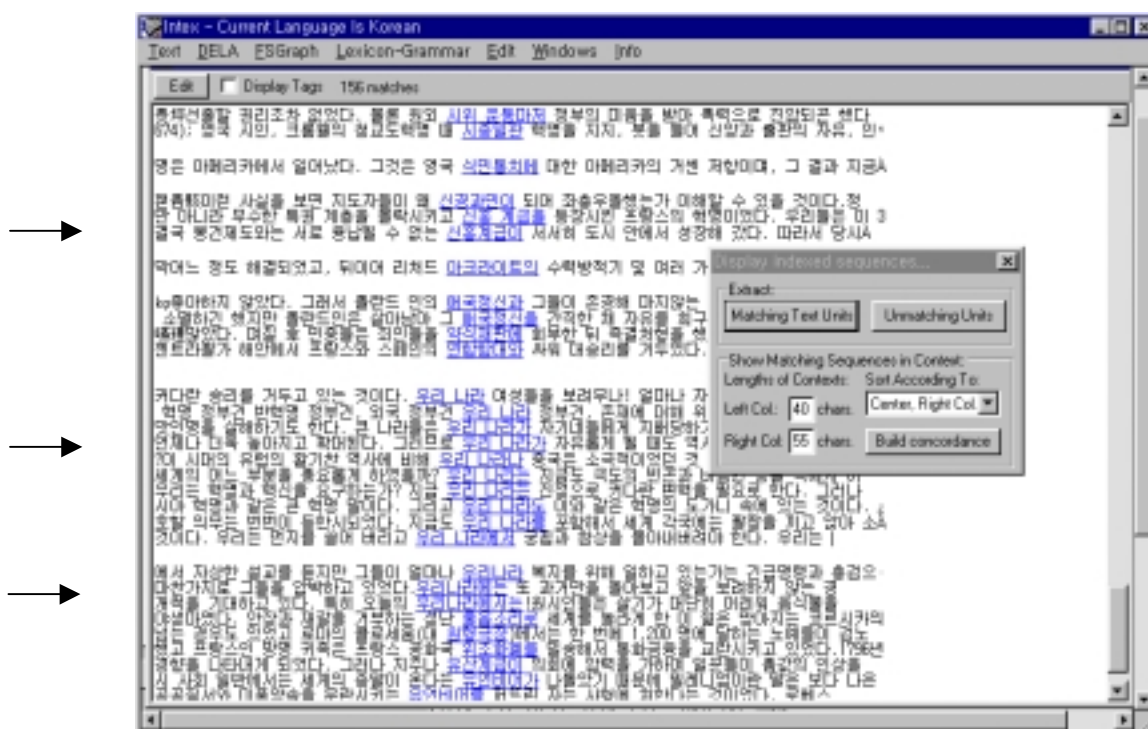


Figure 14. Concordance des noms figés du texte trouvés dans le dictionnaire¹⁴

Dans la figure 14, nous vérifions que l’auteur alterne fréquemment l’utilisation de l’espace pour un même NFF dans son texte. Par exemple, les noms figés *sinheung jyegeup* « ascension - classe = classe ascendante » ou *uli nala* « nous - pays = la Corée » se réalisent différemment dans le même texte (voir →).

Notre expérience (cf. figure 15) de reconnaissance automatique des NFF avec DELAF et DELACF dans plusieurs types de textes montre bien l’importance des variations liées à l’espacement en coréen. Par exemple, les auteurs des textes 4, 5, 6 et 9 ont tendance à insérer l’espace dans les séquences nominales figées, tandis que les auteurs des autres textes privilégient la soudure. Pour traiter les textes coréens, il est donc indispensable de prendre en considération les variantes liées à l’espacement et de préciser les espacements possibles dans le dictionnaire électronique des séquences nominales en coréen.

¹⁴ Pour le moment, INTEX sous Windows coréen peut afficher le coréen pour plusieurs fonctionnalités, mais ne peut pas afficher directement la concordance. Cependant, nous pouvons afficher indirectement le coréen dans INTEX grâce aux fonctions de « copier » et « coller » du logiciel Microsoft Word.

Type du texte ¹⁵	Taille du texte	Nombre de NFF reconnus à la main	Nombre de NFF reconnus par DELAF (A)	Nombre de NFF reconnus par DELACF (B)	Nombre de NFF reconnus (A + B)	Portion reconnue par DELAF et DELACF
1 : Essai C	100 Ko	624	172	90	262	42 %
2 : Histoire N	100 Ko	270	154	15	169	63 %
3 : Société H	100 Ko	842	267	7	274	33 %
4 : Journal D1	100 Ko	462	7	270	277	60 %
5 : Journal D2	100 Ko	535	8	265	273	51 %
6 : Journal CH	100 Ko	414	16	192	208	50 %
7 : magazine K	100 Ko	736	169	9	178	24 %
8 : Journal HK	100 Ko	881	199	53	252	29 %
9 : Journal HKL1	100 Ko	515	3	386	389	76 %
10 : Journal HKL2	100 Ko	696	130	66	196	28 %
Total	1 Mo	5975	1125	1353	2478	41 %

Figure 15. Portion reconnue des NFF par DELAF et DELACF¹⁶

Dans notre expérience, nous avons trouvé 5975 NFF à la main dans les 10 textes (soit un total de 1 Mo). Dans les 10 textes, 2478 NFF sur 5975 NFF sont reconnus par la consultation de nos dictionnaires (soit 41% de NFF). Par conséquent, nous pouvons estimer en moyenne que 6 NFF apparaissent par Ko de texte et nos dictionnaires DELAF et DELACF reconnaissent 2,5 NFF par Ko de texte. Nous pouvons augmenter ce nombre par ajout des noms figés non reconnus dans notre dictionnaire électronique.

2-2. Perspectives pour le traitement des séquences nominales non reconnues

Il existe trois raisons pour expliquer la non-reconnaissance de certaines séquences nominales : la première raison vient de l'insuffisance du dictionnaire, la deuxième des séquences nominales libres soudées et la dernière des noms prédicatifs combinés avec les suffixes verbaux comme *-hada* « faire » ou des noms combinés avec le copule *-ida* « être ». Nous discutons de ces trois problèmes dans le traitement des

¹⁵ cf. Pour la source, voir la Bibliographie.

¹⁶ Nous avons choisi 10 textes au hasard dans le corpus du *Projet Sejong 21*.

séquences nominales.

Par notre expérience, nous constatons que notre dictionnaire de 45000 NFF n'est pas suffisant pour reconnaître tous les NFF dans les textes coréens. Pour compléter le dictionnaire des séquences nominales figées, il est nécessaire d'étendre les dictionnaires des NFF à l'aide de corpus et il sera indispensable de construire les dictionnaires de NC et de NFO après des études systématiques sur les autres types de noms (noms compacts dérivés, noms figés à espacement obligatoire). En outre, il est nécessaire au moins de construire les dictionnaires des noms propres figés (par exemple, les toponymes, les noms de pays, etc.) et les noms techniques figés en considérant leur espacement dans l'usage, bien que nous ne puissions pas mettre tous les termes néologiques dans un dictionnaire électronique. Pour le traitement automatique des séquences nominales, nous avons besoin de 7 types de dictionnaires électroniques des séquences nominales :

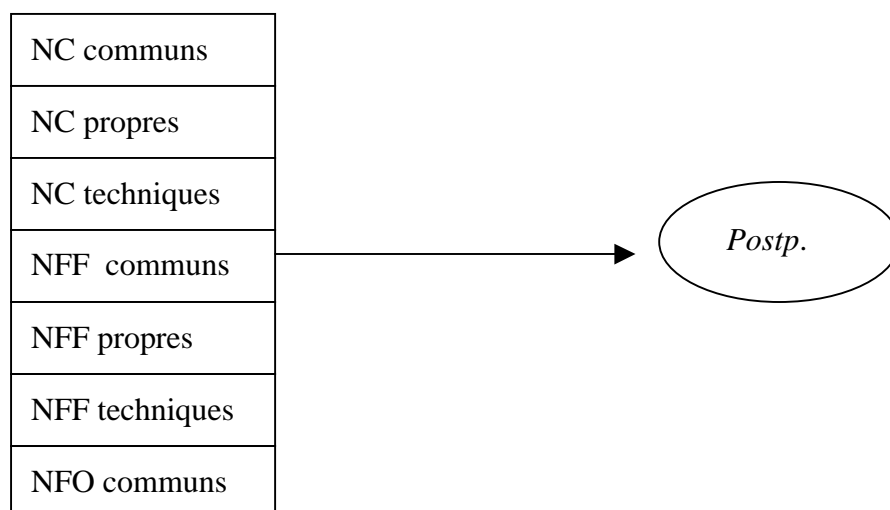


Figure 16. Dictionnaires des séquences nominales¹⁷

Or, bien que nos dictionnaires soient assez importants en nombre d'entrées, nous sommes confrontés à un autre type de problème lors du traitement automatique des séquences nominales, à savoir la reconnaissance des séquences nominales libres soudées. Ces deux problèmes, la reconnaissance des séquences nominales figées et la reconnaissance des séquences nominales libres soudées, sont liés aux conventions typographiques. Comme nous l'avons vu dans notre expérience (cf. figure 15), les

¹⁷ Il nous semble qu'il existe peu de NFO propres et de NFO techniques.

variantes d'espacement des séquences nominales figées peuvent être identifiées par la construction des dictionnaires des séquences nominales figées qui précisent les variations typographiques. Cependant, le problème des séquences libres soudées constituées de plusieurs noms n'est pas résolu par la construction d'un dictionnaire électronique le plus exhaustif possible. Ce problème est capital dans la mesure où les applications pratiques comme la traduction automatique ou l'extraction de mots-clés demandent une analyse précise des séquences nominales. Pour limiter le nombre des mots non reconnus, il faut segmenter les séquences nominales libres soudées en appliquant les dictionnaires des noms et des séquences nominales figées. Pour obtenir une segmentation précise des séquences nominales libres, a priori, il faut construire un dictionnaire électronique suffisamment exhaustif des séquences nominales figées, car une séquence nominale libre peut contenir une séquence nominale figée comme nous l'avons vu dans le chapitre 2.

Même si l'on segmente les séquences nominales libres soudées en appliquant des dictionnaires de séquences nominales figées suffisamment exhaustifs, on peut trouver des cas de segmentations ambiguës¹⁸. Nous ne pouvons pas entrer dans les détails, mais nous remarquons le problème des segmentations ambiguës des séquences nominales libres :

(3) a. *dae-hak-saeng-seon-gyo-hoe* (étudiant - réunion de la mission = réunion de la mission des étudiants)

b. *daehaksaeng / seongyohoe*

(étudiant / réunion de la mission = réunion de la mission des étudiants)

c. *daehak / saengseon / gyohoe* (université / poisson / église = ?)

(Yoon B. H et al., 1995)

Par exemple, la séquence libre (3a) peut être segmentée de plusieurs façons sans dictionnaire exhaustif. Avec un dictionnaire suffisamment complet, il existe encore deux possibilités : (3b) et (3c). Le découpage correct est (3b), car (3c) n'a aucun sens en coréen. Pour obtenir la segmentation exacte, il serait indispensable d'en savoir plus sur la grammaire des séquences nominales libres à espacement facultatif.

La troisième raison de l'existence de noms figés inconnus vient du cas où un nom prédicatif est combiné avec un suffixe verbal comme *-hada* « faire » ou du cas où

¹⁸ Pour les solutions statistiques, voir la section 2-2 dans le chapitre 1.

un nom est combiné avec la copule *-ida* « être »¹⁹. Il s'agit de noms prédicatifs qui entrent dans la structure $N_0 \quad W \quad N_{pred-hada}$ et de noms qui entrent dans la structure $N_0 \quad W \quad N-ida$. Si le but est d'extraire tous les noms figés, il faudrait ajouter les formes fléchies de *Npred-hada* et de *N-ida* dans notre dictionnaire des formes fléchies des NFF après des études systématiques sur les suffixes flexionnels des verbes.

3. Avantages de notre dictionnaire électronique pour plusieurs applications du traitement automatique

Nous explorons les avantages de notre dictionnaire électronique pour plusieurs applications du traitement automatique comme la recherche d'informations, l'extraction de mots-clés, la traduction automatique, etc.

3-1. Recherche d'informations et extraction de mots-clés

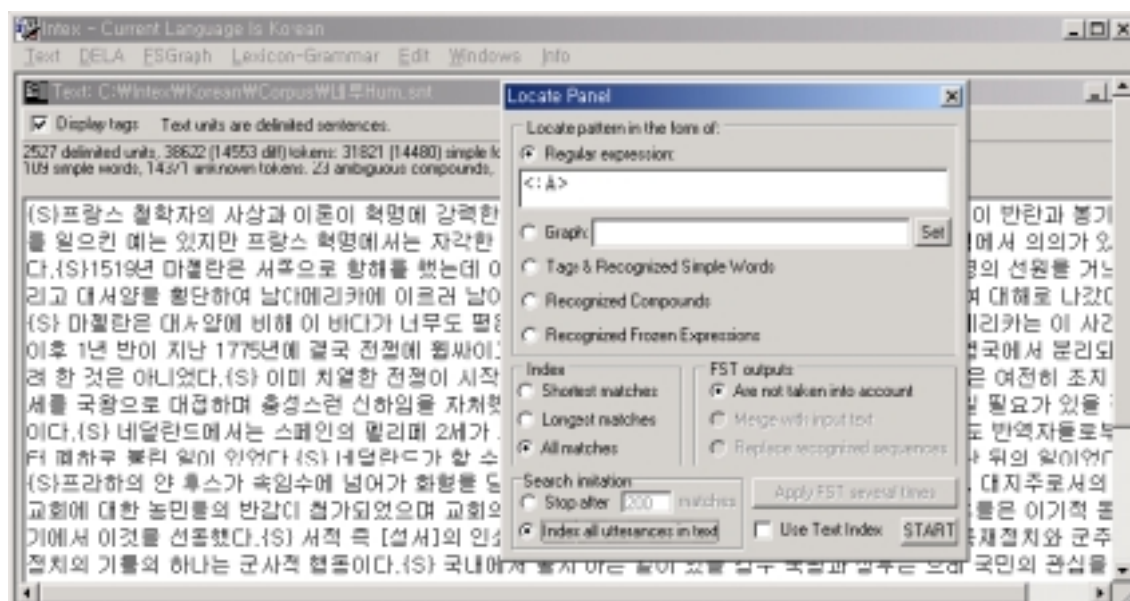
Les systèmes coréens les plus représentatifs comme l'analyseur morphologique de Kang S.S (1993 et 1998), celui de *KAIST*, celui du *Projet Sejong 21*, etc. ne considèrent pas les variantes liées à l'espacement dans le dictionnaire, et ils segmentent les formes fléchies des séquences nominales présentes dans les textes coréens en une séquence nominale et une postposition nominale avec un dictionnaire de noms et un autre de postpositions nominales sans fusionner ces deux ensembles de données. En cas de segmentation ambiguë, chaque système utilise une technique spécifique : ces cas exceptionnels sont répertoriés à part dans un dictionnaire de cas exceptionnels, ils sont résolus par apprentissage automatique à partir de gros corpus, etc.

Lorsque nous utilisons notre dictionnaire des formes fléchies des NFF dans le domaine de la recherche d'information, le résultat de la recherche est beaucoup plus précis que celui de systèmes préexistants grâce à notre méthodologie de construction du dictionnaire. En vue de la reconnaissance des séquences nominales figées par dictionnaire, nous avons classé celles-ci en trois catégories selon les conventions typographiques : noms compacts, noms figés à espacement facultatif et noms figés à espacement obligatoire, puis nous avons séparément construit le dictionnaire électronique de 45000 NFF et celui des séquences de postpositions nominales et enfin nous fusionnons ces deux ensembles de données. De plus, nous avons décrit des informations linguistiques utiles dans le dictionnaire des NFF.

¹⁹ En coréen, tous les noms peuvent se combiner avec la copule *-ida* « être ». C'est pourquoi selon certaines linguistes, la forme *-ida* est considérée comme une postposition nominale, dite prédicative.

yeolu hwaga, *yeolu hwaga*.N31+Hum
yeolu hwagaga, *yeolu hwaga*.N31+Hum: *ga***Nmtf**

Nous pouvons rechercher n'importe quelle postposition nominale ou une séquence de postpositions nominales donnée. Par exemple, nous pouvons trouver les séquences nominales qui contiennent les variantes *-ga/-i* de la postposition nominale du nominatif par l'expression régulière $\langle :A \rangle^{20}$. Remarquons que nous avons codé *A* la postposition du nominatif $\langle ga \rangle$ dans le chapitre 4 :



Nous constatons que notre dictionnaire reconnaît bien les séquences nominales figées qui contiennent les variantes du nominatif *<ga>*.

166

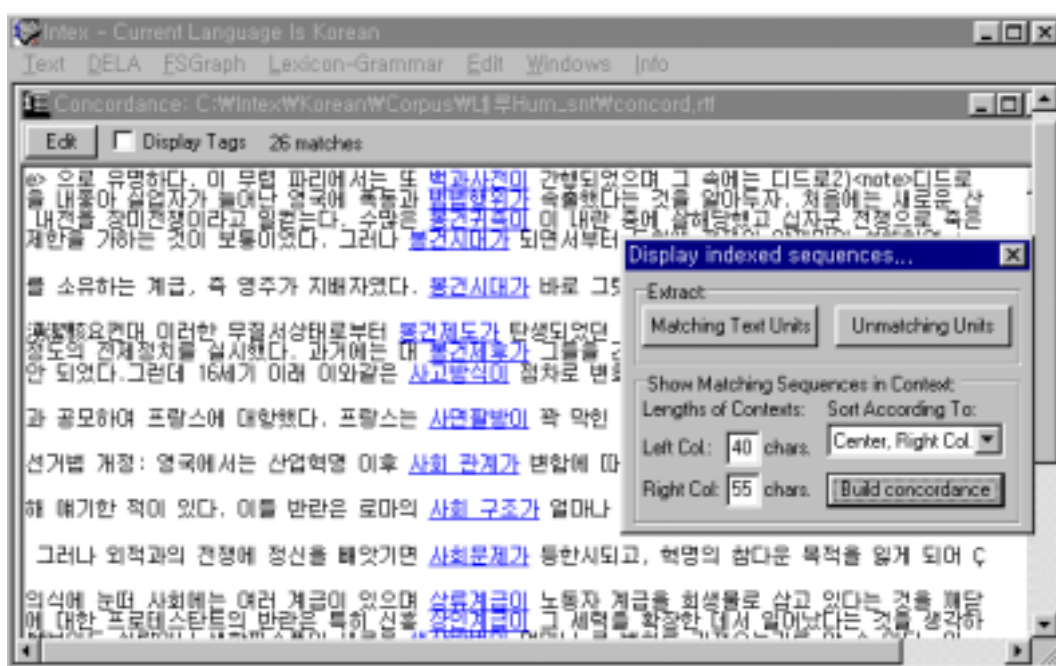


Figure 18. Concordance de la postposition <ga> dans un texte

Les autres systèmes ne permettent pas la recherche des séquences qui contiennent une certaine postposition en n'importe quelle position dans les séquences parce que le dictionnaire de postpositions nominales n'est qu'une liste de postpositions nominales (ou de séquences de postpositions nominales) et les séquences de postpositions nominales n'y sont pas segmentées en postpositions nominales. Mais notre dictionnaire le permet : si nous voulons rechercher toutes les séquences (*N-jocha*, *N-eseo-jocha*, *N-eseo-jocha-do*, etc.) qui contiennent la postposition *-jocha* en n'importe quelle position dans les séquences, nous pouvons les rechercher par l'expression régulière <:3> où 3 est le code de la forme canonique et la classe linguistique de la postposition nominale *-jocha* (<*jocha*, **Aux**>).

Les informations sur le nom tête de chaque séquence sont évidemment utiles pour l'indexation en vue de la recherche d'informations ou l'extraction de mots-clés. La fonction de sous-classifieur de N_1 dans N_1N_2 ou de X dans XN est observée dans le chapitre 2 : les critères N_1N_2 est un ($N_2 + \text{type de } N_2$) et XN est un ($N + \text{type de } N$) sont appliqués pour l'identification des noms figés. Nous avons codé cette information (cf. chapitre 3) et rappelons ci-dessous un exemple pour chaque type (a) - (h) :

- (4) a. *daehap jogae*, $N+NN+12$ « palourde » « coquillage » = « palourde »
- b. *iut sachon*, $N+NN+1$ « voisin » « cousin » = « voisin plus familier que des

cousins »

- c. *chuli soseol*, $N+NN+2$ « déduction » « roman » = « roman policier »
- d. *jigak daejang*, $N+NN+0$ « retard » « chef » = « personne qui est toujours en retard »
- e. *minju jeongchi*, $N+XN+2$ « démocratique » « politique » = « politique démocratique »
- f. *gwagyeok bunja*, $N+XN+0$ « radical » « élément » = « extrémiste ou radical »
- g. *beseuteu seleo*, $N+XX+0$ « best » « seller » = « succès de librairie »
- h. *bitamin di*, $N+NX+1$ « vitamine » « D » = « vitamine D »

Par exemple, lors de l'indexation ou de l'extraction de mots-clés, le code +12 permet d'indexer la séquence figée entière ainsi que chaque nom N_1 et N_2 . Le code +1 (ou +2) permet d'indexer la séquence figée entière et le nom N_1 (ou N_2). Le code +0 empêche l'indexation des 2 éléments et autorise seulement l'indexation de la séquence figée entière. Par exemple :

- (5) a. *daehap jogae*, $N+NN+12$: *daehap jogae*, *daehap*, *jogae*
b. *iut sachon*, $N+NN+1$: *iut sachon*, *iut* (**sachon*)
c. *chuli soseol*, $N+NN+2$: *chuli soseol*, *soseol* (**chuli*)
d. *jigak daejang*, $N+NN+0$: *jigak daejang* (**jigak*, **daejang*)
e. *minju jeongchi*, $N+XN+1$: *minju jeongchi*, *jeongchi* (**minju*)
f. *gwagyeok bunja*, $N+XN+0$: *gwagyeok bunja* (**gwagyeok*, **bunja*)
g. *beseuteu seleo*, $N+XX+0$: *beseuteu sele* (**beseuteu*, **seleo*)
h. *bitamin di*, $N+NX+1$: *bitamin di*, *bitamin* (**di*)

3-2. Analyse syntaxique

Les codes indiquant la catégorie grammaticale, les informations sémantiques, les informations morpho-syntaxiques et syntaxiques sont indispensables non seulement pour l'analyse morphologique mais aussi pour l'analyse syntaxique.

Nous avons codé les noms humains *Hum*, les noms d'animaux *Anm*, les noms de plantes *Plt*, les noms institutionnels humains *InstHum* et les noms humains collectifs *CollHum*. L'information sémantique joue un rôle très important pour la description syntaxique du lexique-grammaire (cf. figure 19).

Les noms prédicatifs (soit compacts soit figés) qui fonctionnent comme prédicats sémantiques dans les phrases, requièrent une description séparée de leurs

comportements syntaxiques dans le lexique-grammaire. Nous avons codé les noms figés prédicatifs dont le nom tête est morpho-syntaxiquement associé à une construction verbale par la transformation « PREDHN=V-n ». Ce code permet de réutiliser les tables syntaxiques du lexique-grammaire des noms compacts prédicatifs²¹. Nous avons tiré un exemple de la description des noms prédicatifs à verbe support *hada*. Par exemple, *gyeolhon* « mariage » dans la table WNH2SYM²² est un nom compact prédicatif et *gukje gyeolhon* « international » « mariage » = « mariage avec un étranger » est un nom figé prédicatif dont le nom tête *gyeolhon* « mariage » est morpho-syntaxiquement associé à une construction verbale. *gukje gyeolhon* partage le comportement de *gyeolhon* et nous pouvons donc réutiliser cette table du lexique grammaire.

²¹ Han, S.H (2000). ANNEXE.

²² Han, S.H (2000). ANNEXE, p. 72.

3-3. Traduction automatique

Notre dictionnaire peut être étendu en un dictionnaire bilingue pour la traduction automatique des séquences nominales figées. Par exemple, la forme canonique associée à chaque entrée du dictionnaire de NFF (DELAC) peut être utilisée dans des applications de traduction.

(6) a. *gukchaek eunhaeng*, banque nationale. N12+NN+sk_sk-PRED+2+InstHum

b. *gukchaek eunhaeng*, government owned bank.

N12+NN+sk_sk-PRED+2+InstHum

(6a) est pour la traduction du coréen vers le français et (6b) pour celle du coréen vers l'anglais. Les transducteurs ne peuvent pas être directement appliqués en mode de remplacement pour les noms coréens, parce qu'en général les noms apparaissent sous une forme fléchie dans les textes. Il faut prendre en considération les postpositions nominales attachées aux noms. Pour la traduction automatique des noms coréens, tout d'abord, il faut appliquer le dictionnaire des formes fléchies des séquences nominales figées. Cette étape permet de segmenter les postpositions nominales. Ensuite, les transducteurs des séquences nominales figées doivent être appliqués en mode remplacement.

Quant à la traduction des postpositions nominales qui est plus complexe que celle des séquences nominales figées, elle est différente selon les classes de postpositions nominales. Par exemple, les postpositions auxiliaires qui ajoutent une nuance à la phrase comme *-do* « aussi », *-man* « seulement », etc. peuvent être décrites dans les cas simples dans les transducteurs de traduction, mais les postpositions nominales dites adverbiales comme *-e* « à », *-eseo* « de », etc. requièrent l'utilisation du lexique-grammaire des prédicats parce que leurs emplois sont liés au comportement syntaxique des prédicats. Les postpositions du nominatif ou de l'accusatif donnent une fonction syntaxique au substantif dans la phrase du coréen, et dans la langue cible la fonction du substantif dans la phrase n'est pas nécessairement la même. Il faut étudier chaque postposition nominale pour la traduction automatique.

On pourrait écrire des transducteurs de traduction pour les séquences nominales, mais il est difficile de les utiliser directement pour plusieurs raisons : les variantes liées à l'espacement, la segmentation entre les postpositions nominales et les séquences nominales, etc.

4. Conclusion

En vue de la reconnaissance automatique des NFF, nous avons séparément construit le dictionnaire électronique de 45000 NFF et celui des séquences de postpositions nominales, et puis nous avons fusionné ces deux ensembles de données. Nous récapitulons les avantages principaux de notre dictionnaire des formes fléchies des NFF (DELAF pour les NFF sans espace et DELACF pour NFF avec au moins un espace) :

- (i) Il permet de rechercher toutes les variantes de NFF liées à l'espacement
- (ii) Il permet de segmenter précisément les formes fléchies des NFF en un NFF et une séquence de postpositions nominales
- (iii) Il permet de segmenter précisément les séquences de postpositions nominales

De plus, il comporte les codes utiles pour plusieurs applications du traitement automatique comme la recherche d'informations, l'extraction de mots-clés, la traduction automatique, etc. :

- (i) Code indiquant un trait sémantique et le statut de nom prédicatif
- (ii) Code indiquant le nom tête de chaque entrée
- (iii) Code indiquant l'origine et la catégorie grammaticale

CONCLUSION

Dans ce travail, nous avons présenté des méthodes originales de construction de dictionnaires électroniques de séquences nominales figées et de leurs formes fléchies, et nous avons justifié la validité de nos méthodes en appliquant notre dictionnaire.

En vue de la reconnaissance des séquences nominales figées par dictionnaire, nous avons classé celles-ci en trois catégories selon les conventions typographiques : noms compacts (NC), noms figés à espacement facultatif (NFF) et noms figés à espacement obligatoire (NFO). Nous avons construit un dictionnaire électronique des NFF à 45000 entrées. Puisqu'en général, des formes fléchies des séquences nominales figées apparaissent dans les textes coréens, en vue de la constitution de dictionnaires exploitables des séquences nominales figées, nous avons séparément construit le dictionnaire des séquences de postpositions nominales avec leur segmentation, sous forme d'un transducteur fini, et enfin nous avons fusionné ces deux ensembles de données à partir de codes flexionnels du dictionnaire électronique et de la fonctionnalité de flexion d'INTEX. Le nombre de formes fléchies de NFF dans notre dictionnaire est de 200×10^6 (environ 12 Go) et sa taille comprimée est d'environ 1 Go. Dans notre expérience, nous avons trouvé 6 NFF par Ko de texte à la main, et 2,5 NFF par Ko de texte (soit 41 % de NFF) sont reconnus par la consultation de notre dictionnaire des formes fléchies des NFF. Nous pouvons augmenter le nombre reconnu par ajout des noms figés inconnus dans notre dictionnaire électronique.

Notre dictionnaire construit d'après ces méthodes a les principaux avantages suivants par rapport aux systèmes préexistants :

- (i) DELAF et DELACF de NFF permettent la reconnaissance automatique de toutes les variantes de NFF liées à l'espacement
- (ii) DELAF et DELACF de NFF permettent la segmentation des formes fléchies des NFF en un NFF et une séquence de postpositions nominales sans erreur
- (iii) Le dictionnaire des séquences de postpositions nominales sous forme de graphes permet leur segmentation en postpositions nominales sans erreur
- (iv) Le dictionnaire des NFF sert à la segmentation des séquences nominales libres soudées
- (v) Le dictionnaire des NFF peut être étendu en un dictionnaire bilingue pour la traduction automatique

De plus, chaque entrée du dictionnaire de NFF comporte des codes utiles pour

les applications dans le traitement automatique :

- (i) Code indiquant un trait sémantique et le statut de nom prédicatif
- (ii) Code indiquant le nom tête de chaque entrée
- (iii) Code indiquant l'origine et la catégorie grammaticale

Nous constatons que ces codes linguistiques sont très utiles pour plusieurs applications dans le traitement automatique comme la recherche d'informations, l'extraction de mots-clés (ii) et l'analyse syntaxique ((i) et (iii)).

Notre dictionnaire des NFF à 45000 entrées n'est pas suffisant pour reconnaître toutes les séquences nominales figées communes. Pour compléter le dictionnaire des séquences nominales figées et de leurs formes fléchies, il est nécessaire d'étudier les autres types de noms (noms compacts dérivés, noms figés à espacement obligatoire) et d'étendre les dictionnaires de séquences nominales figées à l'aide de corpus. Par l'étude de la combinatoire syntactico-sémantique entre noms et postpositions nominales, il faut améliorer la classification des séquences nominales figées.

En informatique, le problème de la segmentation correcte des séquences nominales non enregistrées dans un dictionnaire est toujours ouvert : si les groupes nominaux libres sont soudés, ils peuvent par définition être absents du dictionnaire électronique même complet. Pour réduire le nombre des séquences nominales non reconnues, il faut obligatoirement appliquer une procédure de segmentation à l'aide, notamment, des dictionnaires de séquences nominales figées. La segmentation correcte des séquences soudées du type de NN^+ du coréen est un processus indispensable pour diverses applications comme la traduction automatique, la correction d'espacement, l'indexation automatique, l'extraction de mots-clés, etc. Bien que le dictionnaire électronique soit suffisamment exhaustif, on peut trouver des segmentations ambiguës. Pour la segmentation précise de ces cas, il est indispensable d'étudier la grammaire des séquences nominales libres à espacement facultatif.

BIBLIOGRAPHIE

Dictionnaires (Le numéro dans le texte est donné entre parenthèses)

- Hangeul Hakhoe (Société pour Hangeul). 1991. *Ulimal Keun Sajeon* (Grand Dictionnaire du Coréen) (Dic 1)
- Unpyeong Eomun Yeonguso. 1997. *Keumsungpan Gukeo Daesajeon* (Grand Dictionnaire du Coréen dans la version de Keum-sung Publ.Co). Keum-sung Publ.Co. (Dic 2)
- Yonsei Eoneojeongbo Yeonguwon (Centre de recherches sur le coréen de l'Université Yonsei) 1998. *Yonsei Hangukeo Sajeon* (Yonsei Dictionnaire du Coréen). Séoul : Doosan Dong-A Publ.Co. (Dic 3)
- Guklip Gukeo Yeonguwon (l'Académie Nationale de la Langue Coréen). *Pyojun Gukeo Daesajeon* (Grand Dictionnaire Standard du Coréen). 1999. Séoul : Doosan Dong-A. (Dic 4)
- Lee, Sung-gu. 1995. *Ttueosseugi Silmu Sajeon* (Practical Spacing Dictionary). Séoul : Aepeul Gihwoek. (Dic 5)
- Hong, Chai-Song et al. 1997. *Hyeondae Hangukeo Dongsa Gumun Sajeon* (Dictionnaire des constructions verbales du coréen contemporain), Séoul : Doosan Dong-A Publ. Co.

Corpus utilisé pour l'expérience

(CD-ROM de corpus écrit du Projet Sejong 21 (<http://www.sejong.or.kr>))

- Texte 1** (Essai). Chun Gu-Seok (1993). *I ttangdeongi-wa bapsang* (Cette terre et la table à manger). Séoul : Changjak-gwa Bipyeongsa.
- Texte 2** (Histoire). J. Nehru (1993) (traduction par Seoktap Press). *Segyesa* (Histoire du monde). Séoul : Seoktap Press.
- Texte 3** (Société ; culture de masse). Hanul Academy (1990). *Hyeondae sahoe-wa maeseukeomunikeisyeon* (La société moderne et la communication de masse)
- Texte 4** (Editoriaux). Journal de Dong-A 1 (1990).
- Texte 5** (Editoriaux). Journal de Dong-A 2 (1990).
- Texte 6** (Editoriaux). Journal de Chosun (1992).
- Texte 7** (Magazine politique). Gyeonghyang sinmun (Presse de journal de Gyeonghyang) (1993). *Nyuseu pipeul* (News People).
- Texte 8** (Sections de science). Journal de Hanguk (1996).

Texte 9 (Sections d'économie). Journal de Hangyeolye 1 (1991).

Texte 10 (Sections de culture). Journal de Hangyeolye 1 (1992).

OUVRAGES ET ARTICLES

- Bae, Sun-Mee. 1999. 1. Typologie des noms composés en coréen 2. Classification et combinatoire des suffixes nominaux en coréen. *Mémoire de DEA*, IGM, Université de Marne-la-Vallée.
- Bae, Sun-Mee. 2000. Vers le traitement des groupes nominaux du coréen dans INTEX. *Revue. Informatique et Statistique dans les Sciences Humaines*, vol. 36, pp. 9-28.
- Bae, Sun-Mee. 2001a. Description of frozen nominal sequences in Korean in an electronic dictionary. *Proceedings of 4th Intex workshop*, Bordeaux, June 2001. (<http://www.nyu.edu/pages/linguistics/intex>)
- Bae, Sun-Mee. 2001b. Construction of an electronic dictionary for compound nouns in Korean. *2nd International Conference ASIALEX. Proceedings*, Seoul, August 8-10, 2001, pp. 182-187.
- Bae, Sun-Mee. 2001c. Deux types de représentations des séquences nominales figées en coréen et de leurs formes fléchies avec INTEX. *Lingvisticae Investigationes* 24:2, Amsterdam / Philadelphia : John Benjamins B.V., pp. 147-165 (apparaître).
- Bae, Sun-Mee. 2002. Construction of an electronic dictionary for compound nouns in Korean. *Sajeon Pyeonchanhak Yeongu* (Studies in Lexicography), je 11 jip 1 ho (Tome 11: 1). Séoul : Centre de recherches sur le coréen de l'Université Yonsei, pp. 151-164.
- Baptista, Jorge. 1995. Estabelecimento e formalização de classes de nomes compostos. *Tese de Mestrado*, Faculdade de Letras da Universidade de Lisboa.
- Choi, Jae-Hyuk. 1996. A Division Method of Korean Compound Noun. *je 8 hoe Hangeul mit Hangukeo jeongbocheoli haksulbalpyo nonmunjip* (8th proceedings for Korean and Korean information), pp. 262-267.
- Choi, Sung-Woo. 1996. Hangul Code Manager for KSC5601 – 1987 WANSUNG Code. *Rapport Technique* N° 31, IGM. Université de Marne-la-Vallée.
- Choi, Sung-Woo. 1999. Implantation de dictionnaires électroniques du coréen par automates finis. *Thèse de doctorat*, IGM. Université de Marne-la-Vallée.
- Chrobot, Agata. 2001. Recensement et description des mots composés – méthodes et applications. *Thèse de doctorat*, LADL. Université de Marne-la-Vallée.
- Christine, Michaux. 1992. The collectives in French : a linguistic investigation. *Lingvisticae Investigationes* XVI : 1, Amsterdam / Philadelphia : John

- Benjamins B.V., pp. 99-124.
- Chung, Min-Chung. 1998. Les nominalizations d'adjectifs en coréen. Constructions nominales à support ISSDA. *Thèse de doctorat*. Paris : Université de Paris 7.
- Courtois, Blandine. 1990. Un système de dictionnaires électroniques pour les mots simples du français. *Langue française* N° 87, Paris : Larousse, pp. 11-22.
- Danlos, Laurence. 1992. Support verb constructions : linguistic properties, representation, translation. *Journal of French Language Studies*, Vol. 2, Cambridge University Press, pp. 1-32.
- Daciuk, Jan. et al. 2000. Incremental Construction of Minimal Acyclic Finite State Automata, *Computational Linguistics* 26 (1), pp. 3-16, MIT Press.
- Dubois et al., 1973. *Dictionnaire de linguistique*. Paris : Larousse.
- Gross, Gaston et Vivès, Robert. Les constructions nominales et l'élaboration d'un lexique-grammaire. *Langue Française* N° 69, Paris : Larousse, pp. 5-27.
- Gross, Gaston. 1986a. Typologie des noms composés : le lexique électronique des noms composés du français, *Rapport ATP*, CNRS, Paris : Université de Paris 13.
- Gross, Gaston. 1989. *Les constructions converses du français*. Genève : Droz.
- Gross, Gaston. 1990a. Définition des noms composés dans un lexique-grammaire. *Langue Française* N° 87, Paris : Larousse, pp. 84-90.
- Gross, Gaston. 1990b. Degré de figement des noms composés. *Langages* N° 90, Paris : Larousse, pp. 57-72.
- Gross, Gaston. 1990c. Les mots composés. *Modèles linguistiques* Tome XII : 1, pp. 47-63.
- Gross, Gaston. 1995. A propos de la notion d'humain. *Linguisticae Investigationes*, J. Labelle et Ch. Leclère (eds.), LIS 17, Amsterdam / Philadelphia : John Benjamins B.V., pp. 71-80.
- Gross, Gaston. 1996. *Les expressions figées en français. Noms composés et autres locutions*. Paris : OPHRYS.
- Gross, Maurice. 1975. *Méthodes en syntaxe*, Paris : Hermann.
- Gross, Maurice. 1981. Les bases empiriques de la notion de prédicat sémantique, *Langages* N° 63, Paris : Larousse, pp. 7-52.
- Gross, Maurice. 1986a. Les nominalizations d'expressions figées, *Langue Française* N° 69, Paris : Larousse, pp. 64-84.
- Gross, Maurice. 1986b. The representation of compound words. *11th International Conference on Computational Linguistics (COLING'86)*.
- Gross, Maurice. 1990. Les limites de la phrase figée. *Langages* N° 90, Paris : Larousse, pp. 7-22.

- Gross, Maurice. 1994. Dictionnaire électronique et traduction automatique. *Langages* N° 116, Paris : Larousse.
- Gross, Maurice. 1997a. The construction of local grammars dans *Finite-State Language Prosessing*. MIT press.
- Gross, Maurice. 1997b. Synonymie, morphologie dérivationnelle et transformations. *Langages* N° 128, Paris : Larousse, pp. 72-90.
- Giry-Schneider, Jacqueline. 1978. *Les nominalisations en français*. Genève : Droz.
- Giry-Schneider, Jacqueline. 1986. Les noms construits avec faire : compléments ou prédicats ?, *Langue Française* N° 69, Paris : Larousse, pp. 49-63.
- Giry-Schneider, Jacqueline. 1987. *Les prédicats nominaux en français. Les phrases simples à verbe support*. Genève : Droz.
- Han, Sun-Hae. 1995. Problème de description des noms prédicatifs en coréen. Sur le verbe *nata* (se produire). *Mémoires du CERIL* N° 13, IGM, Université de Marne-la-Vallée, pp. 38-69.
- Han, Sun-Hae. 1996. La formation des noms en *-jil* en coréen. *Papers in Computational Lexicography, COMPLEX*, Budapest : Research Institute for Linguistics, Hungarian Academy of Sciences, pp. 96 - 105.
- Han, Sun-Hae. 2000. Les prédicats nominaux en coréen – Constructions à verbe support HATA. *Thèse de doctorat*. Paris : Université de paris 7.
- Heo, Ung. 1983. *Gukeohak : ulimal-ui oneul.eoje* (Linguistique coréenne : Hier et aujourd'hui). Séoul : Saem Munhwasa.
- Heo, Ung. 1995. *20 Segi ulimal-ui hyeongtaelon* (La morphologie du coréen pour 20^{ième} siècle). Séoul : Saem Munhwasa.
- Hong, Chai-Sung. 1992. Dongsae meokda-ui sajeonjeok cheoli-leul wihan myeot gaji nonui (Quelques considerations pour une description lexicographique du verbe meokda (manger)). *Sae gukeo saenghwal* Vol 2, N° 4, Séoul : Gukeo Yeonguso, pp. 14-37.
- Hong, Chai-Sung. 1993. Yaksok-ui munbeop : seousulmyeongsa-ui eohwi tongsajeok gisul-gwa sajeon (Grammaire de *Yaksok* (promesse). Une description lexicosyntaxique des noms prédicatifs et dictionnaires). *Dongbanghakji* Vol. 81, Séoul : Université Yonsei, pp. 213-252.
- Hong, Chai-Sung. 1997. Idong dongsae-wa gineung dongsae (Le verbe du mouvement et le verbe support). *Mal* (Le mot) N° 22, Séoul : Université Yonsei, pp. 121-140.
- Hong, Chai-Sung. 1999. Gineung dongsae gumun yeongu-ui han sigak : eohwijeok jeopgeun (Un point de vue sur les etudes du verbe support : l'approche lexicale). *Inmunnonchong*, N° 41, Séoul : Université nationale de Séoul, pp. 135-173.

- Hong, Chai-Sung et al. 1998 ~ 2001. Rapports du Projet Sejong 21. (<http://www.sejong.or.kr>)
- Hopcroft, J. et Ullman, J. (1979). *Introduction to Automata Theory. Languages and Computation*, Adison-Wesley Publishing Company, Reading, Massachusetts.
- Jung, Won-Su. 1992. *Gukeo-ui daneo hyeongseonglon* (La formation des mots du coréen). Séoul : Hansin Munhwasa.
- Jung, Hui-Jung. 2000. *Hangukeo myeongsa yeongu* (Etudes sur les noms coréens). Séoul : Hanguk Munhwasa.
- Jun, Jae-Yeon. 1999. Recherches sur le nombre en coréen : Syntaxe, Sémantique et Morphologie, *Thèse de doctorat*. Université de Paris 8.
- Kang, Seung-Shik. 1993. Korean Morphological Analysis Using Syllable Information and Multi-word Unit Information. *Thèse de doctorat*. Séoul : Université nationale de Séoul.
- Kang, Seung-Shik. 1998. A Decomposition Algorithm of Korean Compound Nouns. *The Korean Information Science Society : Journal of KISS*, Vol. 25-1. pp. 172-182.
- Kuh, Hakan. 1987. Plural Copying in Korean, *Harvard Studies in Korean Linguistics II*. Massachusetts : Harvard University, Cambridge.
- Kim, Chang-Seop. 1996. Gukeo pasaeng-ui tongsalon-jeok munjedeul (Problème syntaxique sur les mots dérivés du coréen). *Lee Gi-Mun gyosu jeongnyeontoimginyeom nonchong* (Publication en commémoration pour Monsieur le professeur Lee Gi-Mun).
- Kim, Gwang-Hae. 1990. Yanghwa pyoheon (Expression sur la quantité) dans *Gukeo tyongu eodikkaji watna* (Où est-on au cours des études sur le coréen ?). Séoul : Dong-A Pubi.Co.
- Kim, Gwang-Hae. 1994. Hapseongeo-lan mueot-inga ? (Qu'est-ce qui est le nom composé ?) *Namjeon Park Gap-Su seonsaeng hwoigap ginyeom nonmunjip* (Publication en commémoration pour Namjeon Park Gap-Su). pp. 41-90. (<http://plaza.snu.ac.kr/~sunnysea/research.htm>)
- Kim, Heung-Kyu et Yoon Ku-Byeong. 1993. *Wae sanyago muleumyeon* (Si on demande pourquoi on vit). Seoul : Hansaem.
- Kim, Jin-Hae. 2000. *Gukeo yeoneo yeongu* (Etudes sur les collocations du coréen). *Thèse de doctorat*. Séoul : Université KyungHee.
- Kim, Seung-gon. 1992. *Hangukeo-ui tossi-wa ssikkeut* (Les postpositions pour les noms, les adjectives et les verbes). Séoul : Seogwang Haksul Jalyosa.
- Kim, Suk-Deuk. 1992. *Ulimal hyeongtaelon* (La morphologie du coréen). Séoul : Tapchulpansa.

- Kim, Young-Hee. 1988. *Hangukeo tongsalon-ui mosaek* (Syntaxe du coréen). Séoul : Tapchulpansa.
- Kolyeo Daehakgyo Daehakguko Pyeonchansil (Editorial Department of Korean for college students in Korea University). 1994. *Eoneo-wa sasang – Eoneo-wa munhwa* (language and thought – language and culture). Séoul : Korea University Press.
- Korean Ministry of Education. 1988. Règles orthographiques du coréen.
- Laporte, Eric. 1990. Reconnaissance des expressions figées lors de l'analyse automatique. *Langages* N° 90. Paris, Larousse, pp 59-70.
- Lee, Gyu-Tae. 1993. *CD-ROM-eulo bo-neun Lee Gyu-Tae koneo* (Editoriaux que Lee Gyu-Tae écrit avec CD-ROM), Séoul : Solbit Midieo.
- Lee, Han-Gyu. 1991. Plural marker Copying in Korean, *Harvard Studies in Korean Linguistics* IV, Harvard University, Cambridge, Messachusetts, pp. 513-528.
- Lee, Hui-Ja et Lee, Jong-Hui. 1998. *Tekseuteu bunseokjeok gukeo josa-ui yeongu* (Etudes sur les postpositions nominales pour l'analyse textuelle). Séoul : Hanguk Munhwasa.
- Lee, Hui-Ja. 1995. Study of combinative relation of Korean idiomatic phrase (hyeondae gukeo gwanyonggu-ui gyeolhapgwangye gochal). *Daedong Munhwa Yeongu*, je 30 jip, daedong munhwa yeonguso, Séoul : Université SungKyunKwan.
- Lee, Kyeong-Sun et al. 2000. Workbench for Constructing Dictionary for Semantic Analysis of Compound Noun, *je 12hwoi Hangeul mit Hangukeo jeongbocheoli haksulbalpyo nonmunjip* (12th proceedings for Korean and Korean information). pp. 149-155.
- Lee Ik-Seop. 1968. *Hanjaeo joeobeop-ui yuhyeong* (Typologie de la formation des mots sino-coréens). Séoul : Eulyu Munhwasa.
- Lee, Sang-Eok. 1993. Gwanyong pyohyeon-gwa hapseongeo-ui bunseok mit eowhibo naeoe-eseoui cheoli (l'expression figée, l'analyse du mot complexe et ses traitement dans le lexique). *Eohakyeongu*, je 29 ho je 3 kwon N° 29, volume 3, Séoul : Université nationale de Séoul.
- Lee, Seung-Gu, 1996. *Gukeo jeongseobeop* (Orthographie du coréen), Séoul : Apeulgihoek.
- Mathieu-Colas, Michel. 1988. Variations graphiques des mots composés dans le PETIT LAROUSSE et le PETIT ROBERT. *Linguisticae Investigationes* XVII : 2, Amsterdam / Philadelphia : John Benjamins B.V., pp. 235-280.
- Mathieu-Colas, Michel. 1989, Typologie des noms composés, *Rapport N° 7 du programme de Recherche Coordonnée « informatique Linguistique »*, Paris :

LADL.

- Mathieu-Colas, Michel. 1990. Orthographe et informatique : établissement d'un dictionnaire électronique des variantes graphiques. *Langue française* N° 87, pp. 104-111.
- Meunier, Annie. 1977. Sur les bases syntaxiques de la morphologie dérivationnelle. *Linguisticae Investigationes* I : 2, Amsterdam / Philadelphia : John Benjamins B.V., pp. 287-332.
- Monceaux, Anne. 1993. La formation des noms composés de structure Nom Adjectif. Elaboration d'un dictionnaire électronique. *Thèse de doctorat*. Paris : Université de Paris 7.
- Monceaux, Anne. 1997. Adjectifs de relation ; complémentation et sous-classification. *Langages* N° 126, pp. 39-59.
- Moon, Seung-Chul. 1995. Plural marker TUL is subject to Principle A, *Harvard Studies in Korean Linguistics* VI, pp. 355-369. Massachusetts : Harvard University, Cambridge.
- Mun, Hyeon-Geum. 1999. *Gukeo-ui gwanyong pyohyen yeongu* (Etudes sur les expressions idiomatiques du coréen), Séoul : Taehaksa.
- Nam, Gi-sim et Go, Yong-geun. 1998. *Pyojun gukeo munbeoplön* (Grammaire Standard du coréen). Séoul : Tapchulpansa.
- Nam, Jee-Sun. 1990. Sur une construction $N_0 N_1$ -ITA en coréen. *Linguisticae Investigationes* XIV : 2, Amsterdam / Philadelphia : John Benjamins B.V., pp. 301-341.
- Nam, Jee-Sun. 1994. Dictionnaire des noms simples du coréen. *Rapport Technique* N° 46, IGM. Université de Marne-la-Vallée.
- Nam, Jee-Sun. 1996. Classification Syntaxique des Constructions Adjectivales en Coréen. Amsterdam / Philadelphia : John Benjamins B.V.
- Nam, Jee-Sun. 1997. *Système électronique de lexiques coréens DECO*. Mémoire présenté en vue de l'obtention du diplôme d'habilitation à diriger des recherches en Linguistique Informatique. IGM, Université de Marne-La-Vallée.
- Nam, Yoon-Jin. 2000. *Hyeondae gukeo-ui josa-e daehan gyelyang eoneohakjeok yeongu* (A Study of Statistical Linguistics on the Nominal Postpositions in Modern Korean). Séoul : Taehaksa.
- No, Myeong-Hui. 1985. Hyeondae gukeo hanjaeo-ui daneogujö yeongu (A Study on the Word Structure of sino-Korean in Modern Korean). *Thèse de doctorat*. Séoul : Université nationale de Séoul.
- Noh, Yun-Chae. 1992. Les constructions converses du coréen. Etudes des prédicats

- nominaux. *Thèse de doctorat*. Paris : Université de Paris 7.
- Park, Eun-Jeong. 1996. *Gukeo daneo hyeongseongbeop yeongu* (Etudes sur la formation des mots du coréen). Séoul : Bagijeong.
- Revuz, Dominique. Dictionnaire et Lexique. Méthodes et algorithmes. *Thèse de doctorat*. Paris : Université de Paris 7.
- Seo, Jeong-Su. 1995. *Gukeo munbeop* (Grammaire du coréen). Séoul : Hanyang University Press.
- Song, Seok-Choong. 1988. A ubiquitous plural marker, *Exploration in Korean Syntax and Semantics*. Institute for East Asian Studies. University of California, Berkeley Center for Korean Studies.
- Silberztein, Max. 1989. Dictionnaires électroniques et reconnaissance lexicale automatique. *Thèse de doctorat*. Paris : Université de Paris 7.
- Silberztein, Max. 1990. Le dictionnaire électronique des mots composés. *Langue française* N° 87, pp. 71-83.
- Silberztein, Max. 1993. Dictionnaires électroniques et analyse automatique de textes. Paris : Masson.
- Silberztein, Max. 1993. Les groupes nominaux productifs et les noms composés lexicalisés. *Linguisticae Investigationes* XVII : 2., Amsterdam / Philadelphia : John Benjamins B.V., pp. 405-425.
- Silberztein, Max, 1996. *INTEX 3.4. Reference Manual*, LADL, Paris : Université Paris 7.
- Silberztein, Max, 1997. The lexical Analysis of Natural Languages dans *Finite-State Language Processing*, pp. 175-203.
- Silberztein, Max. 2000. *Le manuel d'INTEX*. LADL, Paris : Université de Paris 7.
- Yoon, Bo-Hyun et al. 1995. A Method of Korean Compound Noun Analysis using Statistical Information. *The Korean Information Science Society : Journal of KISS*, pp. 925-928.
- Yoon Jun-Tae. 1998. Analysis of Korean Compound Noun using Lexical Information between Nouns. *The Korean Information Science Society : Journal of KISS*, p. 1716-1723. Vol. 25-11.

ANNEXE 1. Définition de l'alphabet du coréen dans INTEX sous Windows 98 ou 2000 en coréen

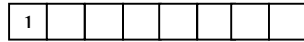
Plusieurs systèmes de code¹ sont développés pour représenter HANGUL (caractères coréens) dans l'ordinateur : par exemple, système de code à N octets, système de code à 2 octets JOHAB, système de code à 2 octets WANSUNG, Unicode, etc. Parmi eux, le logiciel INTEX sous Windows 98 ou 2000 en coréen n'accepte que le système de code *KSC5601-1987 HANGUL WANSUNG* pour l'affichage du coréen. Le *KSC5601-1987 HANGUL WANSUNG*, inventé et sélectionné en 1982 comme système standard par le gouvernement, est un système de code ASCII étendu en 2 octets ; au début, il a été inventé pour l'échange des informations écrites en coréen dans l'ordinateur, mais le gouvernement l'a adapté comme le HANGUL code système dans les réseaux informatiques du gouvernement pour l'administration. Il a été amélioré dans les années 1987 et 1989 par le gouvernement. Maintenant, le *KSC5601-1987 HANGUL WANSUNG* comme système de code standard pour HANGUL est le plus utilisé et répandu pour représenter HANGUL dans l'ordinateur.

Comparons le *KSC5601-1987 HANGUL WANSUNG* avec le système de code ASCII étendu à 8 bits pour représenter les caractères latins accentués. Dans le système de code ASCII étendu à 8 bits pour représenter les caractères latins, la partie supérieure de 0 à 127 (en hexadécimal de 00 à 7F) est réservée pour ASCII caractères, mais la partie inférieure de 128 à 255 (en hexadécimal de 80 à FF) est normalement réservée pour définir les caractères latins accentués. En revanche, dans le système *KSC5601-1987 HANGUL WANSUNG* la partie supérieure est aussi laissée telle quelle pour ASCII caractères, mais la partie inférieure (c'est-à-dire 96 codes de 0xA0 à 0xFF) est réservée pour définir les caractères coréens et chinois. Or, parmi 96 codes, A0 est exceptionnellement réservé pour **Space**, FF pour **DEL** (cf. figure 1). Il ne reste que donc 94 places de A1 à FE. Mais 94 codes ne sont pas suffisants pour définir plusieurs milliers de caractères coréens et chinois utilisés dans la vie quotidienne, à savoir 2350 caractères coréens et 4888 caractères chinois. Pour représenter plusieurs milliers de caractères, on double 94 codes (i.e. $94 * 94 = 8836$). Par conséquent, nous pouvons définir 8836 nouveaux codes de 0xA1A1 à 0xFEFE avec cette méthode d'extension, et ce système comme système de code ASCII étendu en 2 octets satisfait donc la méthode d'extension du code ISO 2022 (International Standards Organization) :

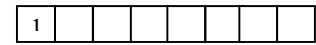
¹ Pour chaque système de code, voir Choi S.W (1996), pp. 5-8.

MSB fixed 1 \Rightarrow Hex $0 \times 80 \sim 0 \times FF$

Upper byte



Lower byte



	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
1	<div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><div><div><div><div></div><div></div></div><div><div></div><div></div></div></div></div><</div></div></div></div></div>															

Combinaison en 2 octets : $94 \times 94 = 8836$ codes possibles

Figure 1. 8 bit code extension method²

Dans la table de code *KSC5601-1987 HANGUL WANSUNG*, les caractères spéciaux et les caractères des langues étrangères comme japonais, allemand, russe, etc. sont assignés selon l'ordre alphabétique dans la zone de 0xA1A1 à 0xACFE, 2350 caractères coréens (c'est-à-dire 2350 syllabes) dans la zone de 0xB0A1 à 0xC8FE et 4888 caractères chinoises³ dans la zone de 0xCA A1 à 0xFDFE. Voici un extrait de la table de code *KSC5601-1987 HANGUL WANSUNG*⁴ :

² Choi, S.W (1996), p. 7.

³ Mathématiquement, en coréen le nombre de toute combinatoire entre les consonnes initiaux (19), les voyelles (21) et les consonnes finals (27) est 11,172 (19 x 21 x 28). Avec Unicode et le système de code à 2 octets JOHAB, on a l'intérêt d'exprimer tous les caractères 11,172. Mais pour le moment ils ne sont pas utilisables dans INTEX.

⁴ Choi, S.W (1996), ANNEXE pp. 1-19.

HEX	DEC	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A1 - A0	161 - 160			、	。	．	…	…	”	”	—	—	∥	＼	～	‘	’
A1 - B0	161 - 176	“	”	[]	<	>	《	》	「	」	『 』	』	【 〓	】	±	×
A1 - C0	161 - 192	÷	≠	≤	≥	∞	∴	∘	′	″	℃	Å	¢	£	¥	♂	♀
A1 - D0	161 - 208	∠	⊥	^	∂	▽	≡	≐	§	※	☆	★	○	●	◎	◇	◆
A1 - E0	161 - 224	□	■	△	▲	▽	▼	→	←	↑	↓	↔	=	≪	≫	√	∞
A1 - F0	161 - 240	∞	∴	∫	∫∫	∈	∋	⊆	⊇	⊂	⊃	⊄	⊅	∧	∨	┐	
A2 - A0	162 - 160		⇒	⇔	∀	∃	∴	∼	∨	∨	”	。	。	，	。	！	？
A2 - B0	162 - 176	:	ℳ	Σ	Π	⊠	°F	‰	◁	◀	▷	▶	♠	♠	♥	♥	♣
A2 - C0	162 - 192	♣	◎	◆	■	●	●	▤	▥	▦	▧	▨	▩	▩	♨	☎	☎
A2 - D0	162 - 208	☎	☎	☎	†	‡	‡	↗	✓	↖	↘	b	J	♪	♪	♫	☎
A2 - E0	162 - 224	No	Co.	TM	am.	pm.	Tel										
A2 - F0	162 - 240																
A3 - A0	163 - 160		!	"	#	\$	%	&	'	()	*	+	,	-	.	/
A3 - B0	163 - 176	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
A3 - C0	163 - 192	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
A3 - D0	163 - 208	P	Q	R	S	T	U	V	W	X	Y	Z	[W]	^	_
A3 - E0	163 - 224	`	a	b	c	d	e	f	g	h	i	j	k	l	M	n	o
A3 - F0	163 - 240	p	q	r	s	t	u	v	w	x	y	z	{		}	—	
A4 - A0	164 - 160		ㄱ	ㄴ	ㄷ	ㄹ	ㅁ	ㅂ	ㅅ	ㅇ	ㅈ	ㅊ	ㅋ	ㅌ	ㅍ	ㅑ	ㅓ
A4 - B0	164 - 176	ㅕ	ㅖ	ㅗ	ㅛ	ㅜ	ㅠ	ㅡ	ㅍ	ㅑ	ㅓ	ㅕ	ㅖ	ㅗ	ㅛ	ㅜ	ㅠ
A4 - C0	164 - 192	ㅑ	ㅓ	ㅕ	ㅖ	ㅗ	ㅛ	ㅜ	ㅠ	ㅡ	ㅍ	ㅑ	ㅓ	ㅕ	ㅖ	ㅗ	ㅛ
A4 - D0	164 - 208	ㅑ	ㅓ	ㅕ	ㅖ	ㅗ	ㅛ	ㅜ	ㅠ	ㅡ	ㅍ	ㅑ	ㅓ	ㅕ	ㅖ	ㅗ	ㅛ
A4 - E0	164 - 224	ㅑ	ㅓ	ㅕ	ㅖ	ㅗ	ㅛ	ㅜ	ㅠ	ㅡ	ㅍ	ㅑ	ㅓ	ㅕ	ㅖ	ㅗ	ㅛ
A4 - F0	164 - 240	ㅑ	ㅓ	ㅕ	ㅖ	ㅗ	ㅛ	ㅜ	ㅠ	ㅡ	ㅍ	ㅑ	ㅓ	ㅕ	ㅖ	ㅗ	ㅛ
A5 - A0	165 - 160		i	ii	iii	iv	v	vi	vii	viii	ix	x	xi	xii			
A5 - B0	165 - 176	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII				
A5 - C0	165 - 192		A	B	Γ	Δ	E	Z	H	Θ	I	K	Λ	M	N	Ξ	O
A5 - D0	165 - 208	Π	P	Σ	T	Υ	Φ	X	Ψ	Ω							
A5 - E0	165 - 224		α	β	γ	δ	ε	ζ	η	Θ	ι	κ	λ	μ	ν	ξ	ο
A5 - F0	165 - 240	π	ρ	σ	τ	υ	φ	χ	ψ	ω							
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
B0 - A0	176 - 160		가	각	간	간	갈	갈	갈	갈	갈	갈	갈	갈	갈	갈	갈
B0 - B0	176 - 176	갈	갈	갈	개	객	객	객	객	객	객	객	객	객	객	객	객
B0 - C0	176 - 192	갓	강	개	객	객	객	객	객	객	객	객	객	객	객	객	객
B0 - D0	176 - 208	갓	갈	갈	갈	갈	갈	갈	갈	갈	갈	갈	갈	갈	갈	갈	갈
B0 - E0	176 - 224	갓	갈	갈	갈	갈	갈	갈	갈	갈	갈	갈	갈	갈	갈	갈	갈
B0 - F0	176 - 240	갓	갈	갈	갈	갈	갈	갈	갈	갈	갈	갈	갈	갈	갈	갈	갈
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
CA - A0	202 - 160		伽	佳	假	價	加	可	呵	哥	嘉	嫁	家	暇	架	枷	柯
CA - B0	202 - 176	歌	珂	痲	稼	苛	茄	街	袈	訶	賈	跏	軻	迦	駕	刻	却
CA - C0	202 - 192	各	恪	慥	殼	珏	脚	覺	角	閣	佩	刊	壑	奸	姦	干	幹
CA - D0	202 - 208	懇	揀	杆	束	桿	澗	癩	看	礪	稈	竿	簡	肝	艮	艱	諫
CA - E0	202 - 224	問	恕	喝	曷	渴	竭	葛	葛	謁	謁	謁	謁	坎	堪	嵌	感
CA - F0	202 - 240	憾	戡	敢	柑	橄	減	甘	疳	監	瞰	紺	邯	鑑	鑒	龕	
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:

Figure 2. Table de code KSC 5601-1987 WANSUNG

Nous avons examiné le système de code *KSC5601-1987 HANGUL WANSUNG* que le logiciel INTEX accepte pour l’affichage du coréen sous Windows 98 ou 2000 en coréen. Dans INTEX, le fichier **Alphabet** est un fichier de type Windows ANSI (ASCII étendu à 8 bits), dans lequel les deux premières lignes du fichier sont destinées au nom et à la taille de polices de caractère utilisées pour afficher des textes et des dictionnaires et ensuite, chaque lettre est recensée et décrite. Or, dans le *KSC5601-1987 HANGUL WANSUNG*, aucun caractère coréen n’est pas défini dans la zone en hexadécimale de A1 à FE (cf. figure 1) et tous les caractères coréens et chinois en 2 octets sont définis dans la zone de A1A1 à FEFE (cf. figure 2). Mais INTEX ne comprend pas le code doublé en 2 octets, et il considère un caractère coréen (ou chinois) comme deux caractères. Dans ce cas, pour la définition d’alphabet du coréen, il est suffisant de déclarer cette zone elle-même en hexadécimale de A1 à FE dans le fichier **Alphabet** dans INTEX⁵. C’est-à-dire qu’il est suffisant de déclarer 94 lignes de A1 à FE (i.e. A1, A2, A3, ...) au lieu de 8836 lignes de A1A1 à FEFE (i.e. A1A1, A1A2, A1A3,...). De plus, puisqu’en coréen il n’y a pas de distinction entre les caractères majuscule et minuscule comme l’anglais ou le français, il est suffisant de doubler le même code de caractère. Le fichier Alphabet serait comme suit :

#"police" taille de police	
#"police" taille de police	
A1A1	} 94 lignes
A2A2	
A3A3	
:	
FDFD	
FEFE	

En se référant à la table de code *KSC 5601-1987 WANSUNG* (cf. figure 2), nous définissons chaque lettre correspondant aux codes de caractère A1A1, A2A2, A3A3, ... , FDFD et FEFE dans le fichier **Alphabet** : par exemple, A1A1 est vide, ⇔ est associé à A2A2, # A3A3, etc. Cet ordre est très important pour bien trier le dictionnaire selon

⁵ Dans le fichier **Alphabet** du système INTEX, il ne s’agit pas de l’ordre alphabétique comme ㄱ, ㄴ, ㄷ, etc. des dictionnaires éditoriaux.

l'ordre alphabétique. Voici le fichier **Alphabet** du coréen dans INTEX :

#"굴림" 11
#"굴림체" 11

⇔

ㄴ
v
ㄴ
ㅁ
L
t
お
カ
K
?
?
?
갈
끗
깨
남
네
도
덤
링
만
뒤
부
뻗
세
슬
씨
옛
응
중
췌
챙
김
텔
파
피
훈

刊
皆
傲
袴
鑛
龜
契
金

菴幢東亮蘆利埋廟縛栢複礪沙賽薛遜旬娠狎憶連霓僚釭幼應壹牆傳程拙葺室寒諦黜拓判瓢桁頰嬋臬羲

ANNEXE 2. Extrait du dictionnaire des NFF

가감^소거법,N13+NN+sk_sk-PRED+2+Inanim
 가감^순서,N33+NN+sk_sk-PRED+2+Inanim
 가감^저항기,N33+NN+sk_sk-PRED+2+Inanim
 가감^축전기,N33+NN+sk_sk-PRED+2+Inanim
 가격^경기,N33+NN+sk_sk+PRED+2+Inanim
 가격^경쟁,N13+NN+sk_sk+PREDHN=Vn+2+Inanim
 가격^기구,N33+NN+sk_sk-PRED+2+Inanim
 가격^변동^준비금,N13+NNN+sk_sk-sk-PRED+3+Inanim
 가격^변동^충당금,N13+NNN+sk_sk-sk-PRED+3+Inanim
 가격^상품권,N13+NN+sk_sk-PRED+2+Inanim
 가격^수준,N13+NN+sk_sk+PRED+2+Inanim
 가격^역지정^주문,N13+NNN+sk_sk-sk+PREDHN=Vn+3+Inanim
 가격^정책,N13+NN+sk_sk+PRED+2+Inanim
 가격^지수,N33+NN+sk_sk-PRED+2+Inanim
 가격^차익금,N13+NN+sk_sk-PRED+2+Inanim
 가격^탄력성,N13+NN+sk_sk+PRED+2+Inanim
 가격^파괴,N33+NN+sk_sk+PREDHN=Vn+2+Inanim
 가격^폭^제한,N13+NNN+sk_sk-sk+PREDHN=Vn+3+Inanim
 가격^표기^우편물,N23+NNN+sk_sk-sk-PRED+3+Inanim
 가격^혁명,N13+NN+sk_sk+PREDHN=Vn+2+Inanim
 가격^효과,N33+NN+sk_sk+PRED+2+Inanim
 가격차^보급금,N13+NN+sk_sk-PRED+2+Inanim
 가계^경제,N33+NN+sk_sk-PRED+2+Inanim
 가계^미가,N33+NN+sk_sk-PRED+2+Inanim
 가계^보험,N13+NN+sk_sk-PRED+2+Inanim
 가계^소득,N13+NN+sk_sk+PRED+2+Inanim
 가계^수표,N33+NN+sk_sk-PRED+2+Inanim
 가계^조사,N33+NN+sk_sk+PREDHN=Vn+2+Inanim
 가계^종합^예금,N13+NNN+sk_sk-sk-PRED+3+Inanim
 가공^형식,N13+NN+sk_sk-PRED+2+Inanim
 가공^경화,N33+NN+sk_sk+PRED+2+Inanim
 가공^도체,N33+NN+sk_sk-PRED+2+Inanim
 가공^무역,N13+NN+sk_sk+PREDHN=Vn+2+Inanim
 가공^배상,N13+NN+sk_sk+PREDHN=Vn+2+Inanim
 가공^삭도,N33+NN+sk_sk-PRED+2+Inanim
 가공^생산,N13+NN+sk_sk+PREDHN=Vn+2+Inanim
 가공^선로,N33+NN+sk_sk-PRED+2+Inanim
 가공^수입,N13+NN+sk_sk+PREDHN=Vn+2+Inanim
 가공^수지,N33+NN+sk_sk-PRED+2+Inanim
 가공^수출,N23+NN+sk_sk+PREDHN=Vn+2+Inanim
 가공^식품,N13+NN+sk_sk-PRED+2+Inanim
 가공^야학금,N13+NN+sk_sk-PRED+2+Inanim
 가공^열처리,N33+NN+sk_sk+PREDHN=Vn+2+Inanim
 가공^원료,N33+NN+sk_sk-PRED+2+Inanim
 가공^유지,N33+NN+sk_sk-PRED+2+Inanim
 가공^의치,N33+NN+sk_sk-PRED+2+Inanim
 가공^이익,N13+NN+sk_sk+PRED+2+Inanim
 가공^이입선,N13+NN+sk_sk-PRED+2+Inanim
 가공^인물,N21+NN+sk_sk-PRED+2+Hum
 가공^자산,N13+NN+sk_sk-PRED+2+Inanim
 가공^전선,N13+NN+sk_sk-PRED+2+Inanim

가공^정원,N13+NN+sk_sk-PRED+2+Inanim
 가공^지선,N13+NN+sk_sk-PRED+2+Inanim
 가공^철도,N33+NN+sk_sk-PRED+2+Inanim
 가공^케이블,N23+NN+sk_bor-PRED+2+Inanim
 가교^결합,N13+NN+sk_sk+PREDHN=Vn+2+Inanim
 가교^엄호,N33+NN+sk_sk+PREDHN=Vn+2+Inanim
 가구^공예,N33+NN+sk_sk-PRED+2+Inanim
 가극^대본,N13+NN+sk_sk-PRED+2+Inanim
 가극^배우,N31+NN+sk_sk-PRED+2+Hum
 가답^재판,N13+NN+sk_sk+PREDHN=Vn+2+Inanim
 가답^판결,N23+NN+sk_sk+PREDHN=Vn+2+Inanim
 가내^공업,N13+NN+sk_sk-PRED+2+Inanim
 가내^공장,N13+NN+sk_sk-PRED+2+Inanim
 가내^노동,N13+NN+sk_sk+PREDHN=Vn+2+Inanim
 가내^노비,N31+NN+sk_sk-PRED+2+Hum
 가내^노예,N31+NN+sk_sk-PRED+2+Hum
 가내^수공업,N13+NN+sk_sk-PRED+2+Inanim
 가능^보조^어간,N13+NNN+sk_sk-sk-PRED+3+Inanim
 가능^수력,N13+NN+sk_sk-PRED+2+Inanim
 가독^상속,N13+NN+sk_sk+PREDHN=Vn+2+Inanim
 가독^상속권,N13+NN+sk_sk-PRED+2+Inanim
 가독^상속인,N11+NN+sk_sk-PRED+2+Hum
 가동^판결,N23+NN+sk_sk-PRED+2+Inanim
 가동^댐,N13+NN+sk_bor-PRED+2+Inanim
 가동^막장,N13+NN+sk_sk-PRED+2+Inanim
 가동^엔제,N33+NN+sk_sk-PRED+2+Inanim
 가동^자본,N13+NN+sk_sk-PRED+2+Inanim
 가동^접점,N13+NN+sk_sk-PRED+2+Inanim
 가동^접촉자,N33+NN+sk_sk-PRED+2+Inanim
 가동^제언,N13+NN+sk_sk-PRED+2+Inanim
 가두^공연,N13+NN+sk_sk+PREDHN=Vn+2+Inanim
 가두^극장,N13+NN+sk_sk-PRED+2+Inanim
 가두^녹음,N13+NN+sk_sk+PREDHN=Vn+2+Inanim
 가두^테모,N33+NN+sk_bor+PREDHN=Vn+2+Inanim
 가두^모금,N13+NN+sk_sk+PREDHN=Vn+2+Inanim
 가두^문학,N13+NN+sk_sk-PRED+2+Inanim
 가두^방송,N13+NN+sk_sk+PREDHN=Vn+2+Inanim
 가두^서명,N13+NN+sk_sk+PREDHN=Vn+2+Inanim
 가두^선교,N33+NN+sk_sk+PREDHN=Vn+2+Inanim
 가두^선전,N13+NN+sk_sk+PREDHN=Vn+2+Inanim
 가두^시위,N33+NN+sk_sk+PREDHN=Vn+2+Inanim
 가두^연극,N13+NN+sk_sk-PRED+2+Inanim
 가두^연설,N23+NN+sk_sk+PREDHN=Vn+2+Inanim
 가두^응원,N13+NN+sk_sk+PREDHN=Vn+2+Inanim
 가두^집회,N33+NN+sk_sk+PREDHN=Vn+2+Inanim
 가두^판매,N33+NN+sk_sk+PREDHN=Vn+2+Inanim
 가두리^양식,N13+NN+pk_sk+PREDHN=Vn+2+Inanim
 가든^골프,N33+XN+bor_bor-PRED+2+Inanim
 가든^파티,N33+XN+bor_bor+PREDHN=Vn+2+Inanim
 가락^국수,N33+NN+pk_pk-PRED+2+Inanim
 가락^단음계,N33+NN+pk_sk-PRED+2+Inanim
 가락^악기,N33+NN+sk_sk-PRED+2+Inanim
 가락^음정,N13+NN+sk_sk-PRED+2+Inanim
 가령^현상,N13+XN+sk_sk-PRED+2+Inanim

ANNEXE 3. Extraits du dictionnaire des séquences de postpositions nominales

